PB83-220111

# Computer Survey of GC/MS (Gas Chromatography/Mass Spectrometry) Data Acquired in EPA's (Environmental Protection Agency) Priority Pollutant Screening Analysis System and Results

(U.S.) Environmental Research Lab., Athens, GA

Jun 83

U.S. Department of Commerce
National Technical Information Service
NTIS

TYPING GUIDE SHEET
BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD, BEGIN SECTIONS HERE

PB83-220111

EPA-600/4-83-029
June 1983

# A COMPUTER SURVEY OF GC/MS DATA ACQUIRED IN EPA'S PRIORITY POLLUTANT SCREENING ANALYSIS: SYSTEM AND RESULTS

by

W. M. Shackelford and D. M. Cline
Environmental Research Laboratory
U.S. Environmental Protection Agency
Athens, Georgia  30613

F. O. Burchfield, L. Faas, G. Kurth, and A. D. Sauter
Computer Science Corporation
Falls Church, Virginia 22046

ENVIRONMENTAL RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
ATHENS, GEORGIA  30613

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"
9-1/8"
PAGE NUMBER
EPA-287 (Cin.)
(4-76)

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE

## TECHNICAL REPORT DATA
*(Please read Instructions on the reverse before completing)*

| | | |
|---|---|---|
| 1. REPORT NO.<br>EPA-600/4-83-029 | 2. | 3. RECIPIENT'S ACCESSION NO.<br>PB83 220111 |
| 4. TITLE AND SUBTITLE<br>A Computer Survey of GC/MS Data Acquired in EPA's Priority Pollutant Screening Analysis: System and Results | | 5. REPORT DATE<br>June 1983 |
| | | 6. PERFORMING ORGANIZATION CODE |
| 7. AUTHOR(S)<br>W.M. Shackelford*, D.M. Cline*, F.O. Burchfield, L. Faas, G. Kurth, and A.D. Sauter | | 8. PERFORMING ORGANIZATION REPORT NO. |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>*Environmental Research Laboratory<br>Athens GA 30613<br>Computer Science Corporation<br>Falls Church VA 22046 | | 10. PROGRAM ELEMENT NO.<br>ABHB1C |
| | | 11. CONTRACT/GRANT NO. |
| 12. SPONSORING AGENCY NAME AND ADDRESS<br>Environmental Research Laboratory--Athens GA<br>Office of Research and Development<br>U.S. Environmental Protection Agency<br>Athens GA 30613 | | 13. TYPE OF REPORT AND PERIOD COVERED |
| | | 14. SPONSORING AGENCY CODE<br>EPA/600/01 |

15. SUPPLEMENTARY NOTES

16. ABSTRACT

The screening analysis phase of the best available treatment (BAT) review of wastewater treatment techniques by EPA was initiated to assess 21 industrial categories for the 129 "priority pollutants." Implicit in the purpose of the screening analysis for these pollutants was the notion that the raw GC/MS data would be saved for later evaluation for compounds not on the priority pollutant list. To this end, a system of computer programs was built that automatically extracts the pure spectra of components in a GC/MS run, matches these spectra against a reference library, and deals appropriately with matched and unmatched spectra. Matched components are entered into a data base for statistical studies to determine their priority for further study. Unmatched spectra are compared to each other to find recurring unknowns so that priorities for ab initio identifications may be set. Component software was obtained from Stanford University (CLEANUP) and Cornell University (PBM) as well as being written at the Athens Environmental Research Laboratory.

17. KEY WORDS AND DOCUMENT ANALYSIS

| a. DESCRIPTORS | b. IDENTIFIERS/OPEN ENDED TERMS | c. COSATI Field/Group |
|---|---|---|
| | | |

| | | |
|---|---|---|
| 18. DISTRIBUTION STATEMENT<br>RELEASE TO PUBLIC | 19. SECURITY CLASS *(This Report)*<br>UNCLASSIFIED | 21. NO OF PAGES<br>185 |
| | 20. SECURITY CLASS *(This page)*<br>UNCLASSIFIED | 22. PRICE |

EPA Form 2220-1 (9-73)

# NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD, BEGIN SECTIONS HERE

# FOREWORD

Nearly every phase of environmental protection depends on a capability to identify and measure specific pollutants in the environment. As part of this Laboratory's research on the occurrence, movement, transformation, impact, and control of environmental contaminants, the Analytical Chemistry Branch identifies and measures chemical constituents of water and soil.

In response to a Consent Decree from the U.S. District Court in the District of Columbia, the U.S. Environmental Protection Agency established a project to assess the wastewaters of 21 industrial categories for 129 Priority Pollutants by gas chromatography/mass spectrometry. Raw GC/MS data from this project were analyzed at the Athens Laboratory to identify additional compounds using a sophisticated, computer-aided system for automatic spectrum extraction and spectra matching. The report describes the project and presents a user's guide for the computerized data survey system.

David W. Duttweiler
Director
Environmental Research Laboratory
Athens, GA

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

PAGE NUMBER
EPA-257 (Cin.)
(4-76)

TYPING GUIDE SHEET

## ABSTRACT

The screening analysis phase of the best available treatment (BAT) review of wastewater treatment techniques by EPA was initiated to assess 21 industrial categories for the 129 "priority pollutants." Implicit in the purpose of the screening analysis for these pollutants was the notion that the raw GC/MS data would be saved for later evaluation for compounds not on the priority pollutant list. To this end, a system of computer programs has been built that automatically extracts the pure spectra of components in a GC/MS run, matches these spectra against a reference library, and deals appropriately with matched and unmatched spectra. Matched components are entered into a data base for statistical studies to determine their priority for further study. Unmatched spectra are compared to each other to find recurring unknowns so that priorities for ab initio identifications may be set. Component software was obtained from Stanford University (CLEANUP) and Cornell University (PBM) as well as being written at the Athens Environmental Research Laboratory.

EPA-287 (Cin.)
(4-76)

# CONTENTS

Foreword . . . . . . . . . . . . . . . . . . . . . . . . iii
Abstract . . . . . . . . . . . . . . . . . . . . . . . . iv
Figures . . . . . . . . . . . . . . . . . . . . . . . . vi
Tables . . . . . . . . . . . . . . . . . . . . . . . . vii
Acknowledgement . . . . . . . . . . . . . . . . . . . . viii

1. Introduction . . . . . . . . . . . . . . . . . . 1
2. Experimental . . . . . . . . . . . . . . . . . . 4
3. Results and Discussion . . . . . . . . . . . . . 13

References . . . . . . . . . . . . . . . . . . . . . . . 32

Appendices

I. GC/MS Data Survey System User's Guide (Version III). . . . . . . . . . . . . . . . 34
II. Applications of a Data Adaptive Background Subtraction Technique to the GC/MS Analysis of Waste Water Samples . . . . . . 150

TYPING GUIDE SHEET
BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

## FIGURES

| Number | | Page |
|---|---|---|
| 1. | Athens EPA computer hardware configuration . . | 6 |
| 2. | Flowchart of software components for run processing . . . . . . . . . . . . . . . . . . | 7 |
| 3. | Logic used in processing a spectrum through HISLIB . . . . . . . . . . . . . . . . . . . | 12 |
| 4. | RGC of 11 component phenol standards. Arrow indicates apparent single component peak that is actually the sum of two components . . | 14 |
| 5. | Resultant spectra from scans 213 and 215 of Figure 4 compared to matching library spectra . . . . . . . . . . . . . . . . . . . | 15 |
| 6. | RGC from an actual extract of an industrial effluent with tentative identifications indicated . . . . . . . . . . . . . . . . . . | 16 |
| 7. | Relationship of identifications made by spectrum matching to the number of components extracted by CLEANUP . . . . . . . | 18 |
| 8. | A. Mixed spectrum extracted by CLEANUP from a VOA standard run. B & C. Library spectra indicated as matched by PBM . . . . . . . . . | 22 |
| 9. | Comparison of spectra for two alkanes. Top spectrum (#226) is tridecane (RRT=.510). Bottom spectrum (#263) is tetradecane (RRT=.591) . . . . . . . . . . . . . . . . . . | 23 |
| 10. | Sample spectrum from MISS library thought to be $D_8$ anthraquinone from degradation of internal standard $D_{10}$ anthracene. This unknown has occurred 11 times in acid extracts run of 1% SP-1240 DA . . . . . . . . | 31 |

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

PAGE NUMBER
EPA-287 (Cin.)
(4-76)

## TABLES

Number Page

I Comparison of Automated vs. Manual Peak Extraction . . . . . . . . . . . . . . . . . . . . 17

II Comparison of PBM Matching Ability for Two Databases . . . . . . . . . . . . . . . . . . . . 20

III Comparison of RRT vs. K Variation for Selected Matched . . . . . . . . . . . . . . . . . . . . 21

IV Mean RRTs for Two Laboratories Before and After Background Correction, Compared with Mean RRTs for Other Laboratories . . . . . . . . . . 24

V Homologous Series Standards Processes for RRT Data . . . . . . . . . . . . . . . . . . . . 25

VI Project Status as of December 31, 1979 Compared with Results to Date . . . . . . . . . . . . . . 25

VII Most Frequently Found Compounds as of December 31, 1979 . . . . . . . . . . . . . . . . 26

VIII Most Frequently Found Compounds as of July 18, 1980 . . . . . . . . . . . . . . . . . . 27

IX Most Frequently Found Compounds as of October 1, 1981 . . . . . . . . . . . . . . . . . 28

X Selected HISLIB Results Indicating "Industry Specific" Compounds . . . . . . . . . . . . . . 29

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 257)

TYPING GUIDE SHEET

# ACKNOWLEDGMENT

We would like to thank Bruce Hodge for his work in bringing CLEANUP up on our computers, Bruce Bartell for writing the graphics software, and David Bradshaw for his work in evaluating data.

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

# SECTION 1

## INTRODUCTION

### BACKGROUND

In June 1976 the U.S. Environmental Protection Agency, as a result of court action by several environmental action groups, was directed by a Consent Decree from the U.S. District Court in the District of Columbia to assess the wastewater of 21 industrial categories for 65 chemical substances and prescribe the best available treatment (BAT) for the effluent. In order to begin the task, a scheme for analysis of the wastewaters for the 65 substances had to be designed.

Although some of the 65 substances were unique chemical compounds, many included whole classes of compounds, e.g. polynuclear aromatic hydrocarbons. Realizing that these classes of compounds could contain literally hundreds of individual members, EPA included for analysis only those members that had been previously identified a significant number of times, were produced in quantity by industry, and were available as analytical standards. The now familiar 129-compound priority pollutant list was the result of this work. Details of the list-resolving procedure may be found elsewhere[1,2].

Even though the list of 129 specific substances made the analysis task manageable, the plaintiffs in the court action were concerned that some members of the chemical classes not on the 129-compound list would be missed in the analysis procedure. Because it was generally agreed that computerized gas chromatography/mass spectrometry (GC/MS) would be the analysis tool of choice, the advantage of saving all raw GC/MS data for later processing to look for compounds other than the priority pollutants became obvious. The state-of-the-art GC/MS instrumentation includes a computer system; thus, the data could be saved in computer readable format for later study. Because magnetic tape is the cheapest mass storage medium, it was chosen for saving all GC/MS data from sample analysis.

Initial analysis of each sample at the laboratories operating under EPA contract was to be directed only toward compounds among the 129 priority pollutants. Although EPA could

BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
BEGIN LAST LINE OF TEXT
3/8"

PAGE NUMBER
EPA-287 (Cin.)
(4-76)

have contracted for a general survey of all compounds in each sample, a number of limiting factors precluded this approach:

A. Cost of a general survey was estimated at $2000 per sample versus $700 for the limited analysis.

B. Time was extremely important. Although the data acquisition times for general survey and specific analysis are the same, data evaluation times could be 5 to 10 times longer for survey analysis (if only computer matching is required for identification). Decreasing the number of samples per unit time by a factor of 5 to 10 would have played havoc with the court-ordered deadlines.

C. Management of the large volumes of unconfirmed data would have required a massive secondary effort to confirm and collate the results of the survey analysis.

By requiring that all data from each GC/MS acquisition be sent to a central location for survey processing, EPA could assure proper management of non-priority pollutant data while at the same time obtain timely response directly from contractor laboratories on the priority pollutants for a reasonable cost. Because all parties involved in the Consent Decree had agreed that the non-priority pollutant data were of less immediate need, no part of the spirit of the Consent Decree was sacrificed; yet provision was made for assessing all the data for compounds other than the 129.

A second provision for possible later analysis of the sample was the requirement for the analysis laboratories to supply each sample extract along with the GC/MS data. Thus, should some compound be tentatively identified in the GC/MS data, it could be confirmed by reanalysis of the corresponding extract. Also, reoccurring components that could not be identified from their mass spectra could possibly be identified using another analysis technique on the saved extract.

The Athens Environmental Research Laboratory had participated in the formation of the priority pollutant list[1] and with the development of the analysis scheme[2]. Responsibility for developing a program for studying the GC/MS data was given to the Athens Laboratory in August 1978. A second program to confirm any tentative mass spectra identifications by reanalysis of the sample extract began in November 1979.

## TASK DESCRIPTION

The screening analysis phase of BAT review was expected to require the qualitative/semi-quantitative analysis of about 4000 samples. Each sample analysis could be expected to

TYPING GUIDE SHEET
EPA-287 (Cin.)
(4-76)

involve GC/MS data acquisition for at least five fractions: a volatile organics analysis (VOA), a VOA blank, an extractable base/neutral (B/N), an extractable acid (ACI), and a direct aqueous injection (DAI). Other blank, standard, and pesticide confirmation runs also could be expected. All calculations for the task were based on 20,000 GC/MS runs, i.e. 4000 samples x 5 fractions. When it is considered that each GC/MS run can be expected to contain some 500 to 1000 individual spectra, the magnitude of the task of evaluating this data is evident.

Implicit in this task was the development of a computer system that could evaluate these data in a manner comparable to a human using computer-aided spectrum extraction and spectra matching to tentatively identify all sample components. An additional goal of the task was to point out those spectra that did not match any spectrum in the reference library yet are seen in multiple GC/MS runs. Thus, a library of compounds tentatively identified in each industrial category as well as a library of recurring but unidentified spectra were to be generated for use in effluent regulation. In addition, the data in these libraries will be studied in a subsequent project in which the saved extracts will be reanalyzed. Tentative identifications made in this project will be confirmed by comparison with standards, and recurring but unidentified spectra will be examined for *ab initio* determination of compound identity.

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN, EPA FORM 267)
TYPING GUIDE SHEET

3

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

## SECTION 2

## EXPERIMENTAL

### SYSTEM DESCRIPTION

The PDP 11/70-based GC/MS Data Survey System consists of computer hardware and software necessary to:

1. Inventory all incoming magnetic tapes and sample extracts.
2. Copy the data on each magnetic tape to a second tape in an internal use format and plot the reconstructed gas chromatogram.
3. Retrieve data as necessary from tapes in batch mode.
4. Extract the spectra of components in each GC/MS run from the background spectra in the run.
5. Match the extracted spectra with a library of reference spectra.
6. Check to see whether matched spectra have been seen before under the same circumstances.
7. Check spectra that are not matched against their fellow unmatched spectra.
8. Generate reports on the numbers of matched spectra by industry, fraction type, analytical laboratory, GC/MS run conditions, etc.
9. Provide graphics capability necessary to view the data from any run.
10. Search any run for specific compounds.

Complete documentation of all software components may be found in Appendix I.

9-1/8"
BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

PAGE NUMBER
EPA-287 (Cin.)
(4-76)

TYPING GUIDE SHEET
BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

## INSTRUMENTATION

Initially, all work for this project used the computer system described in Figure 1. Basically, it consists of a DEC PDP 11/70 with 256K words of memory and 2K words of cache memory, two 88 M-byte disk drives (DEC RP04), two magnetic tape drives (DEC TU16), a Calcomp 936 plotter, and a Tektronics 4014 graphics display terminal. Recently, all run processing functions were transferred to a PDP 11/34 multi-user dedicated system with 128K words of memory and 2K words of cache memory, two 88 M-byte disk drives (DEC RP04), and a Versatec 1200 printer/plotter. This includes spectrum extraction, library matching, and collation of matched and unmatched spectra. A flow chart of these software components is shown in Figure 2. An average of 200 runs per week is processed through this system.

Data submitted to EPA for study was acquired by 18 laboratories using five different systems:

6-1/2"

1. Finnigan 3200 MS, Varian 1400 GC, System Industries data system.
2. Finnigan 4000 MS, Finnigan 9600 GC, Finnigan-Incos data system.
3. Dupont Dimaspec MS, Finnigan 9600 GC, Finnigan-Incos data system.
4. Hewlett-Packard 5985 GC/MS/data system.
5. Hewlett-Packard 5982 GC/MS/data system.

## INVENTORY SYSTEM

In order to inventory and track the 20,000 GC/MS data runs and the estimated 12,000 extracts (a B/N, ACI, and pesticide fraction for each of 4000 samples), a data base management system has been implemented. This system is the INFORM management program and is a well known tool for data base management. In INFORM are kept the GC/MS data run descriptors that allow the physical locating of each run and corresponding extract and also all the information known about the sample.

As each magnetic tape or extract is received at the Athens Laboratory it is manually entered into the INFORM system. Important parameters that are entered for each data run are the tape on which it is found, the EPA sample number, an Athens Laboratory run number, the fraction type, and various GC/MS parameters. The corresponding data for the extract will have all of this information and additionally the precise location of the extract in a freezer.

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA; OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

PAGE NUMBER
EPA-287 (Cin.)
(4-76)

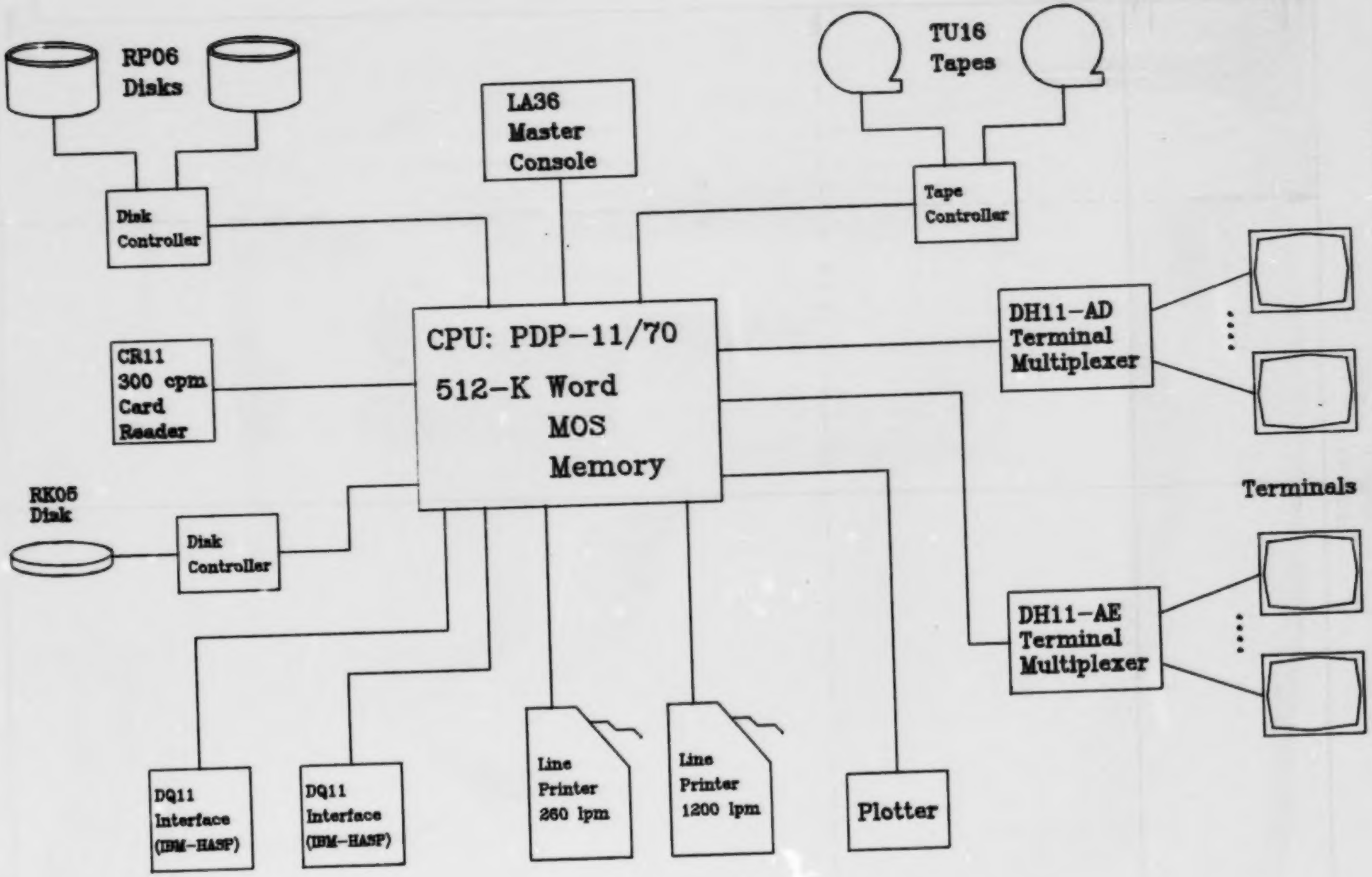

RP06
Disks
LA36
Master
Console
TU16
Tapes
Disk
Controller
Tape
Controller
CPU: PDP-11/70
512-K Word
MOS
Memory
CR11
300 cpm
Card
Reader
DH11-AD
Terminal
Multiplexer
RK05
Disk
Disk
Controller
Terminals
DH11-AE
Terminal
Multiplexer
DQ11
Interface
(IBM-HASP)
DQ11
Interface
(IBM-HASP)
Line
Printer
260 lpm
Line
Printer
1200 lpm
Plotter

Figure 1. Athens EPA computer hardware configuration.

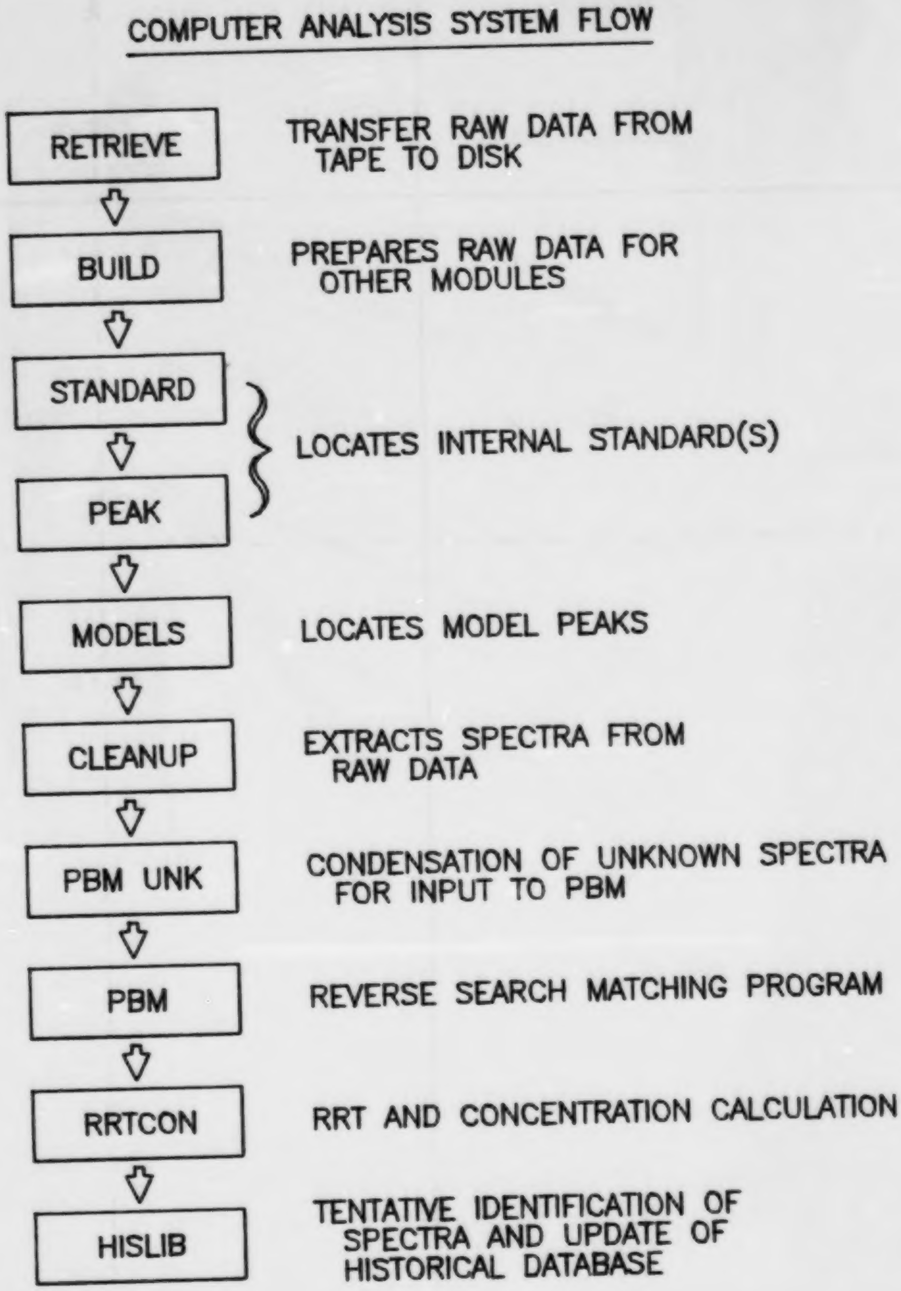

COMPUTER ANALYSIS SYSTEM FLOW
RETRIEVE
TRANSFER RAW DATA FROM TAPE TO DISK
BUILD
PREPARES RAW DATA FOR OTHER MODULES
STANDARD
PEAK
LOCATES INTERNAL STANDARD(S)
MODELS
LOCATES MODEL PEAKS
CLEANUP
EXTRACTS SPECTRA FROM RAW DATA
PBM UNK
CONDENSATION OF UNKNOWN SPECTRA FOR INPUT TO PBM
PBM
REVERSE SEARCH MATCHING PROGRAM
RRTCON
RRT AND CONCENTRATION CALCULATION
HISLIB
TENTATIVE IDENTIFICATION OF SPECTRA AND UPDATE OF HISTORICAL DATABASE

Figure 2. Flowchart of software components for run processing.

During the inventory process, data from the submitting contractor laboratory's tape is copied onto an Athens tape in a format that is both more space efficient and damage resistant. Thus, the original tape and a backup copy are saved. The backup copy, which has only the Athens Laboratory number for identification of each run, is used for all data processing needs. Confidentiality of the data is maintained through the use of the backup copy so that descriptive data is not associated with the GC/MS data. Software that has access to both descriptive and GC/MS data is password protected.

At the time of tape conversion from contractor format to Athens Laboratory format, the data of each run are scanned and a reconstructed gas chromatogram (RGC) is plotted. The RGCs are then bound in volumes to serve as references at the time runs are submitted for analysis. Inspection of the RGC by a chemist may result in discarding the corresponding run because of obvious flaws such as absence of peaks or premature end of data.

When data are to be processed, the chemist identifies the runs that have passed visual inspection for processing. Software then retrieves the designated runs from the magnetic tape and prepares each in turn for processing by the analytical system. The inventory system is reapplied when the run has been processed and descriptors contained in INFORM are necessary for reporting.

## ANALYTICAL SYSTEM

The analytical system consists of four main parts: the internal standard locator, PEAK; the peak or spectrum extractor, CLEANUP; the spectrum matching system, PBM; and the result collator, HISLIB*. Chemists have opportunities at various points during the process to make decisions that can end processing or affect the further processing of any given component spectrum. The ideal data analysis, however, proceeds with minimal operator intervention. Only when the analytical system is presented with decisions that it is unqualified to make does the chemist intervene.

The program PEAK was developed in order to assure the identification of internal standard location in each run.

---

* HISLIB is the name used by Smith et al.[6] to describe their historical library software. The Athens Laboratory-developed historical library uses a similar concept to that of Smith but does not rely on any of Smith's software. CLEANUP and PBM are names of computer programs reported in references (4) and (5), respectively, and are used with slight modification only.

Because all subsequent processing of the data requires knowledge of the internal standard in the run, it was imperative that software be available that would unambiguously define the location and area of the internal standard peak in each data run. The algorithm is discussed in a separate paper[3] that is listed in Appendix II.

CLEANUP is a system of programs developed at Stanford University[4] that finds and extracts the spectra of components in the GC/MS data run. Successive 16-scan windows are searched for ion peaks that have 2 ascending points, a maximum, and 2 descending points. When an ion peak is found, successive ions from mass 40 to 400 are checked to see whether any maximize within a distance of ± 1 scan number of the first found peak. When eight or more such masses maximize simultaneously, a component peak is said to be detected. In this case all the masses maximizing at this point are collected, their areas are normalized to the largest mass of the group, and they are passed along to the next phase of the analysis as a mass spectrum.

CLEANUP involves a number of checks to insure that such artifacts as column bleed, noise spikes, and background are not chosen as sample components. Criteria are input at the start of processing to insure that only ion peaks of a defined sharpness will be considered. This will normally eliminate peaks caused by column bleed, which usually show up in the form of broad peaks. Noise spikes, which are generally of only one-for two-scan duration, are eliminated by requiring a minimum of four scans in the ion peaks. Instrumental background noise caused by pump oil or other contaminants normally does not peak during a run; therefore, it does not interfere with the CLEANUP process.

A preprocessor to CLEANUP called MODELS searches the data for "model" peaks with four ascending and four descending points. Once the model peaks are found, they are used to define the peak shape that a component should have in the region close to the model. A polynomial curve fitting technique is used to redefine the shape of the peak of interest. As one would expect, the model peaks toward the end of the GC/MS run are broader than the models toward the front. Thus, the "real life" conditions of peak shape are taken into account by CLEANUP.

Once a spectrum has been defined by CLEANUP, the area of the base peak is calculated. This is later used to calculate the concentration of the component in the sample by ratioing it with the area of the base peak of the internal standard.

The spectra extracted by CLEANUP are then passed to PBM, a library matching program developed at Cornell University under

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD BEGIN SECTIONS HERE

an EPA grant.[5] PBM, or probability based matching, employs a reverse search technique to compare a reference library of condensed spectra to a similarly condensed unknown spectrum. The algorithm, which is discussed in detail elsewhere[5,7], reports several match parameters by which to evaluate the quality of the matches. An overall match confidence or K value is calculated and compared to $K_{max}$, where $K_{max}$ is the value that can be achieved if all intensities fall within proper windows. These parameters are compared by calculation of K where

$$\Delta K = K_{max} - K \qquad (1)$$

A measure of spectrum contamination is calculated that represents the scaled intensities of ions in the unknown spectrum that are not matched in the reference spectrum and of matched ions in the unknown spectrum that fall outside the intensity tolerances used by PBM. The ratio of the above values to the sum of all scaled intensities for the ions studied (usually the 10 largest ions in the condensed spectrum) is termed "percent contamination." If percent contamination is large, more than one component may be present in the unknown spectrum.

Because the presence of the molecular ion is often helpful in differentiating among the spectra of compounds in a homologous series, PBM notes the presence of a matched molecular ion with a flag (a "+" sign printed with the K value).

A good match, then, has a large K and small $\Delta K$, a low value for percent contamination, and a "+" printed with the K value. In this work, the ratio $K/(K+\Delta K)$ is calculated. Values of this ratio that are greater than 0.45 have been empirically found to indicate reasonable matches.

The 10 best PBM matches for each unknown spectrum are saved and passed to the historical library program. This algorithm looks at the relative retention time (RRT) of each component and the spectrum matches that pass certain match quality criteria ($K/(K + \Delta K) > 0.45$, contamination < 60%) and attempts to match these against entries in the historical library. Each entry in the historical library contains an RRT window, the Chemical Abstracts Service Registry Number (CAS #), and the applicable GC column and internal standard. Therefore, when a match from the spectrum library matches the CAS # and the RRT falls within the window specified in HISLIB, a tentative identification is made. This tentative identification is added to the historical library for reporting purposes.

Of the remaining spectra, those that contain at least one PBM match for which $K/(K + \Delta K) > 0.40$ and percent contamination < 70% are passed to a chemist for inspection, and the rest are

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 287)
TYPING GUIDE SHEET

passed to the MISS library. The results of the chemist's examination may either cause a new entry to be made in the historical library or result in discarding the PBM results. All new entries are confirmed by viewing the spectrum passed by CLEANUP in comparison with library matches suggested by PBM. This is done using a Tektronix 4014 graphics terminal.

These discarded spectra now join those from previous runs that did not match any spectrum in the library of reference spectra. The eight most intense ions for each spectrum are saved and these unmatched spectra are put in the MISS file in order of their RRT. Each spectrum is matched against the spectra in the MISS file that have similar RRT values. If an ion of 75 to 100% relative intensity (RI) is present that agrees with the base peak in the MISS library and six of the remaining ions agree within broad window ranges, a match occurs. In this way unidentified spectra that appear to recur are catalogued. Figure 3 shows the logic used in taking a spectrum through the HISLIB procedure.

## REPORTING

Reporting is accomplished in two ways. The first system is a series of hard copy outputs that describe the flow of data through the total system and the results generated from the data. The contents of the historical library can be printed out either in totality or as a listing of unique entries. The data may be sorted by parameters such as CAS #, RRT, GC column, analysis laboratory, industrial category, relative concentration, etc.

A second method for reporting is a graphics system that allows the chemist to recall data and plot it in various ways. For instance the raw data for a spectrum, the cleaned up spectrum and the reference spectrum can all be plotted on the same screen simultaneously. The extracted ion current profile (EICP) for any ion can be plotted between any scan limits. Multiple EICP plots can be displayed on the screen. The graphics system is used by the chemists to evaluate ambiguous results from the computer analysis.

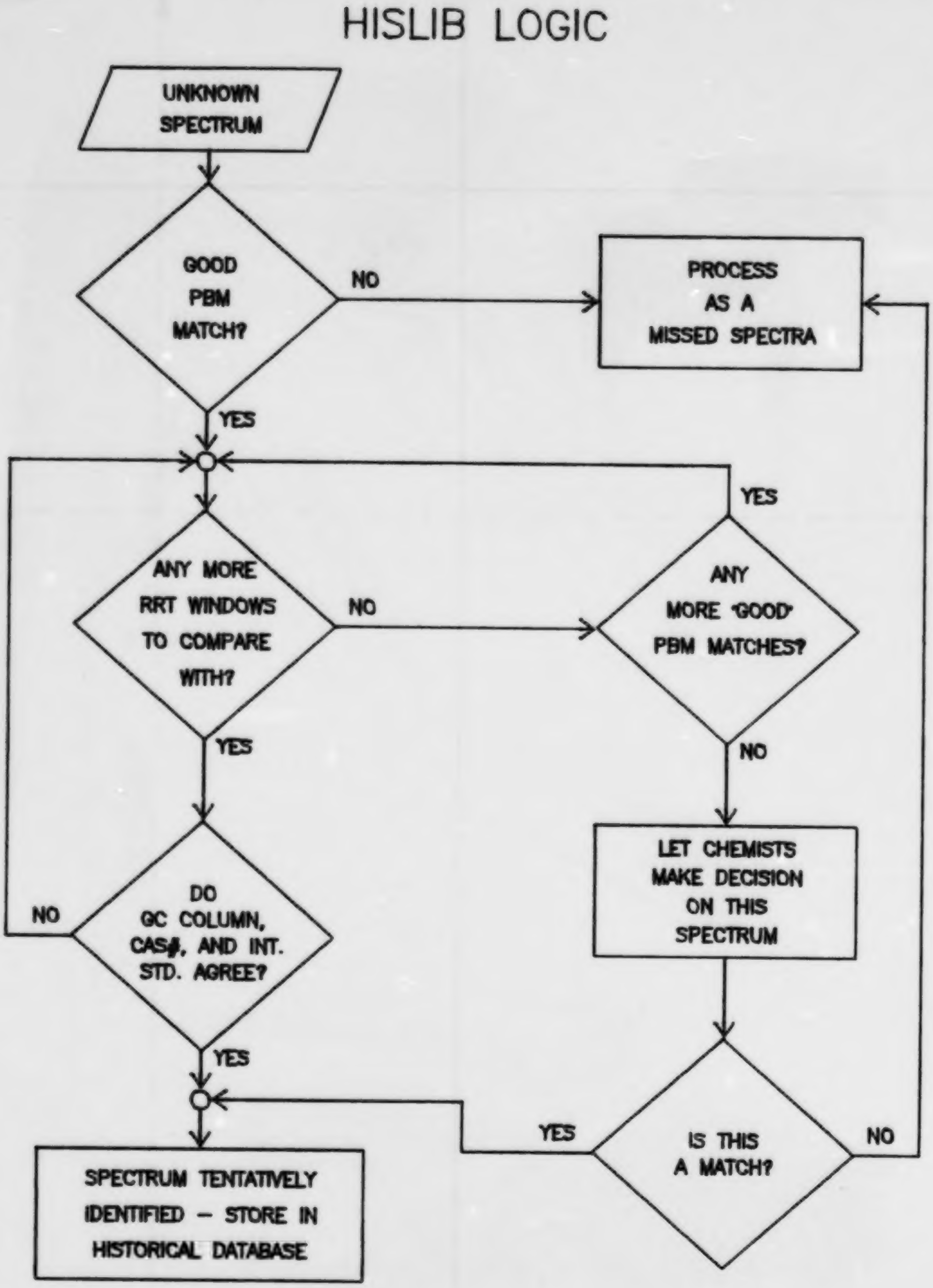

HISLIB LOGIC
UNKNOWN SPECTRUM
GOOD PBM MATCH?
NO
PROCESS AS A MISSED SPECTRA
YES
ANY MORE RRT WINDOWS TO COMPARE WITH?
NO
ANY MORE 'GOOD' PBM MATCHES?
YES
YES
NO
LET CHEMISTS MAKE DECISION ON THIS SPECTRUM
NO
DO GC COLUMN, CAS#, AND INT. STD. AGREE?
YES
IS THIS A MATCH?
YES
NO
SPECTRUM TENTATIVELY IDENTIFIED – STORE IN HISTORICAL DATABASE

Figure 3. Logic used in processing a spectrum through HISLIB.

# SECTION 3

# RESULTS & DISCUSSION

## SPECTRUM EXTRACTION

The extraction of information containing spectra from the mass of spectra in a GC/MS run is the key to a successful automated system for GC/MS data analysis. In Figure 4 an RGC of a group of 11 phenols is shown. The scans 213 and 215 can be seen to be on opposing sides of an apparent single component peak. Manual subtraction of a baseline spectrum (e.g. 208 or 220) from spectrum 214 results in a spectrum that is not recognizable as any of the components injected. Using CLEANUP to find spectra, however, scans 213 and 215 reveal that the peak is actually the sum of two components. Figure 5 shows the resultant spectra of 213 and 215. Also depicted are spectra from the reference library that establish the identity of the two components. Obviously, this example is an ideal case because standards were used with no interferences, but it serves to illustrate the ability of CLEANUP to separate components eluting within two scans of each other.

A second example, Figure 6, shows an actual acid extract from an industrial effluent. Note that spectra were extracted and tentatively identified for 10 components as well as the internal standard, $d_{10}$-anthracene. In Table I, a comparison of component finding ability for manual versus automated techniques is shown. Note that the automated system in each case discerned more components than the manual interpreter. Identification by spectrum matching and manual confirmation are given also. It is interesting to note in these cases that although both manual and automated methods identified the same components (ID's), the automated method, by virtue of more discerned components, identified additional components (ID's in parentheses).

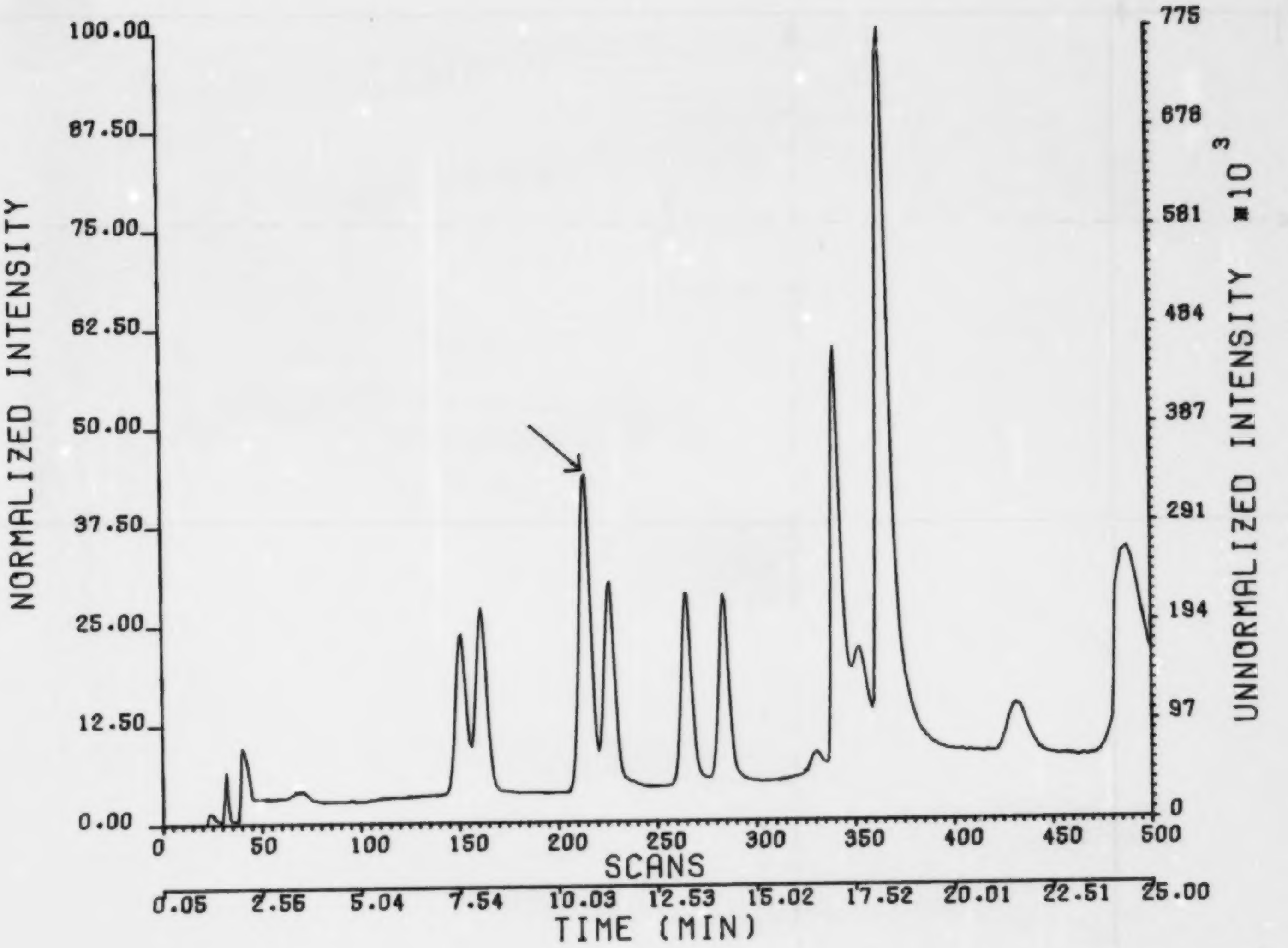


Figure 4. RGC of 11 component phenol standards. Arrow indicates apparent single component peak that is actually the sum of two components.

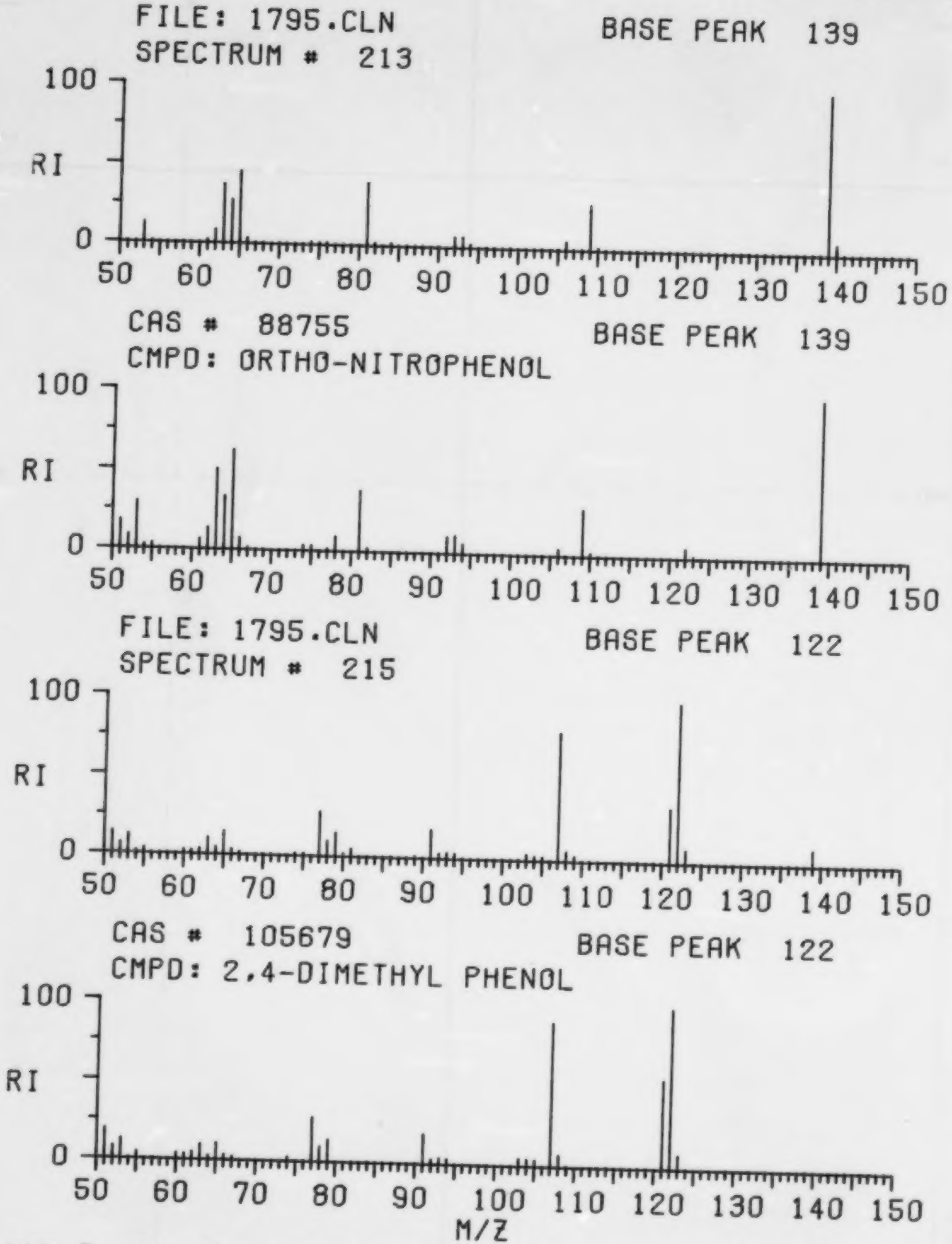


**Figure 5.** Resultant spectra from scans 213 and 215 of Figure 4 compared to matching library spectra.

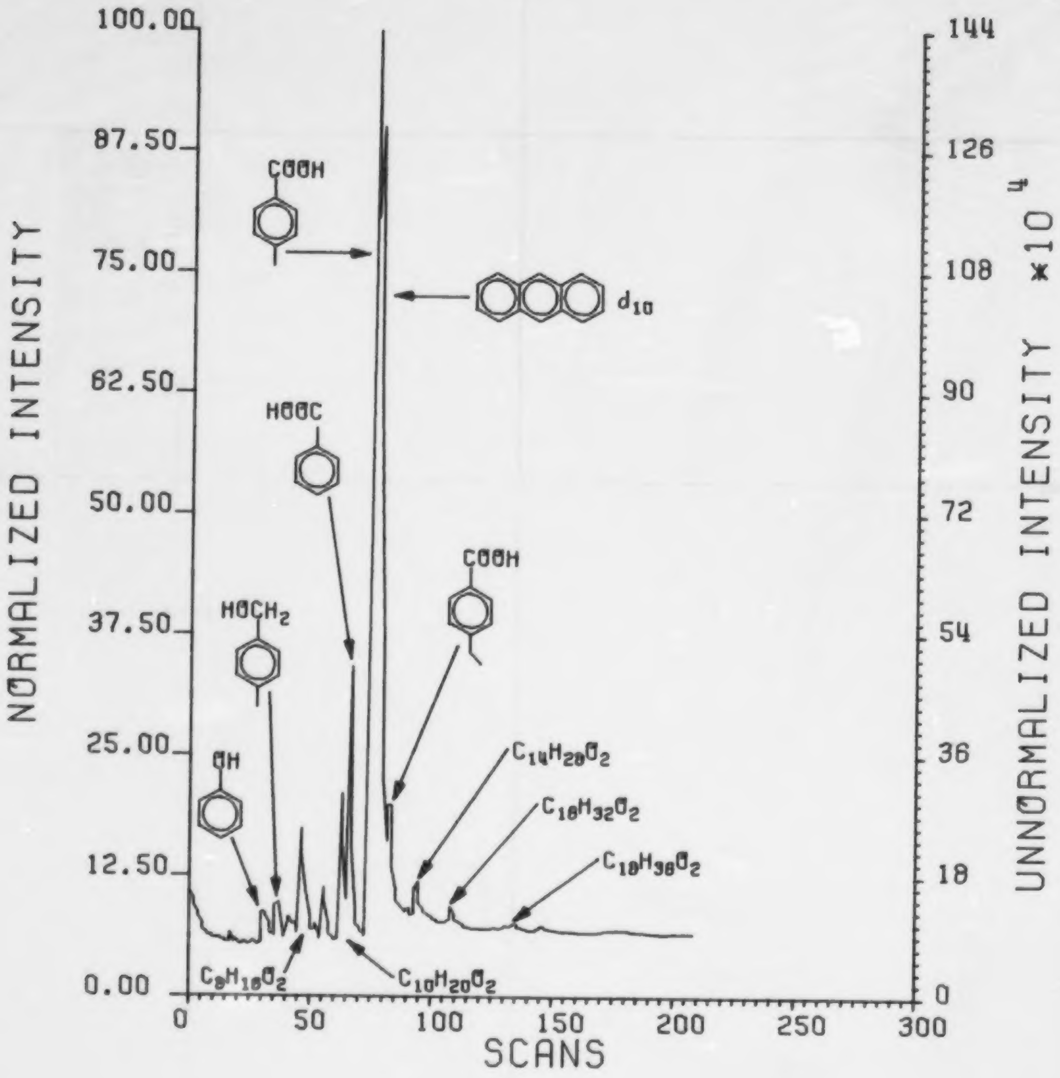


Figure 6. RGC from an actual extract of an industrial effluent with tentative identifications indicated.

TABLE I. COMPARISON OF AUTOMATED VS. MANUAL PEAK EXTRACTION VALUES IN PARENTHESES ARE ADDITIONAL COMPONENTS IDENTIFIED BY THE AUTOMATED METHOD

| Manual | | | CLEANUP-PBM | | |
|---|---|---|---|---|---|
| Peaks | ID's | % ID | Peaks | ID's | % ID |
| 30 | 18 | 0.60 | 46 | 18(6) | 0.52 |
| 18 | 7 | 0.39 | 44 | 7(4) | 0.25 |
| 31 | 11 | 0.35 | 41 | 11(2) | 0.32 |
| 26 | 10 | 0.38 | 43 | 10(2) | 0.28 |

The data shown thus far indicate that for the systems studied, automated techniques are at least equal to manual techniques for pointing out components in the run and identifying them by spectrum matching with a reference library. Before becoming too optimistic, however, it is best to look at cases where spectrum extraction and identification are not so clear cut as in those given above. It can be seen in Table I that despite the fact that more peaks are found with the automated method, the ratio of identifications to peaks has decreased. In fact, as the number of components in a run increases, identification becomes more and more difficult even though the automated system is able apparently to deliver a spectrum for each component. In Figure 7 the relationship of identifications made to components indicated is shown.

One of the reasons for loss in identification efficiency is inherent in CLEANUP. The components that are found that are missed by manual techniques are generally either very small peaks or shoulders on other peaks. CLEANUP then has a difficult time extracting a good spectrum. A good reason is that the general background matrix of components is due to long chain substituted alkyls that are not identified by PBM with any degree of certainty.

CLEANUP requires each potentially detected ion to pass a number of tests before the ion is considered to be detected. As mentioned earlier, each peak must have two rising points, a maximum, and two falling points. The ratio of the slope of these points to the peak width must exceed a user defined value. The area under the described curve must exceed a second user defined value after the local background has been subtracted. These criteria offer no great hurdle for normally chromatographed components with few interferences. Problems develop, however, in the case of more convoluted runs.

The first problem is the loss of weak intensity ions that are important for identification. These ions get lost in the

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA

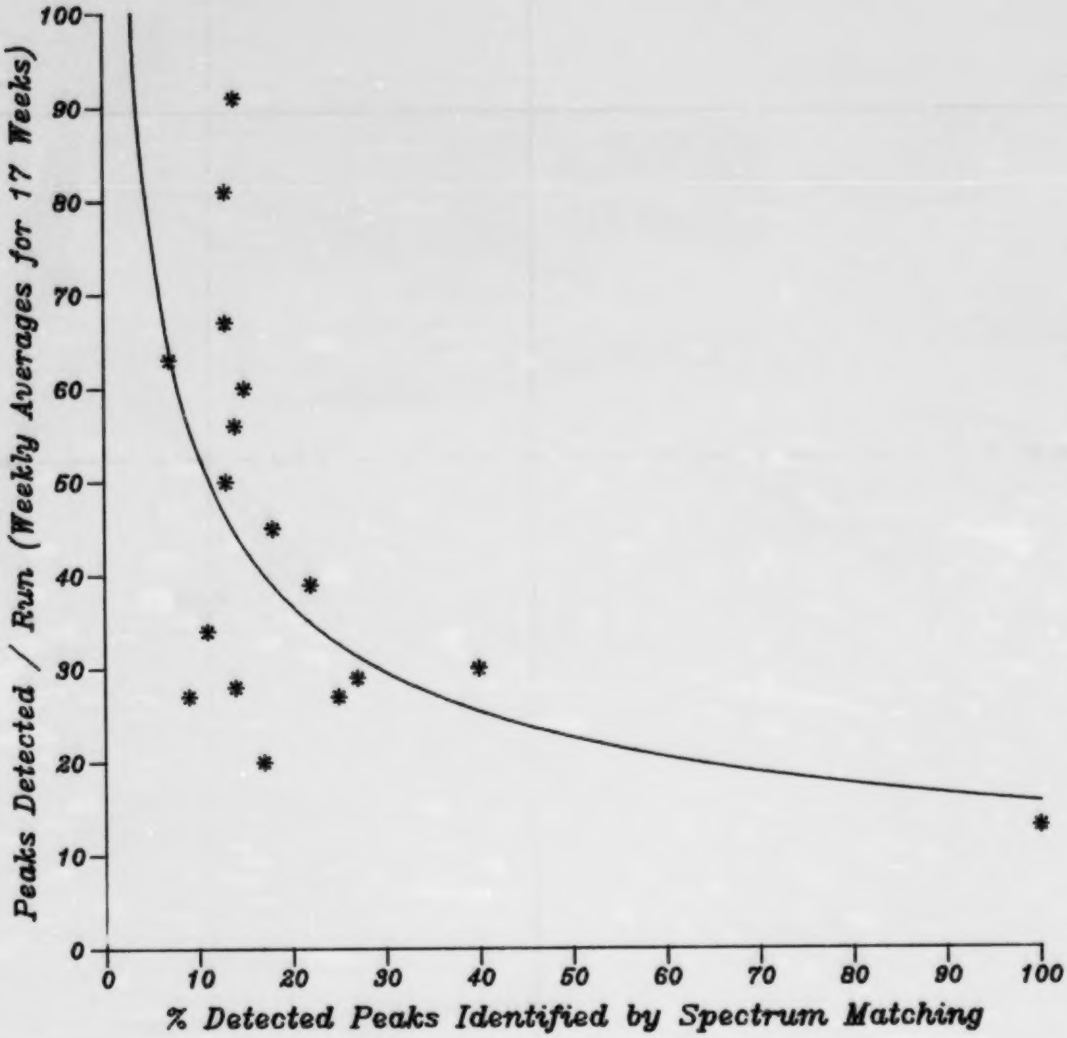


Figure 7. Relationship of identifications made by spectrum matching to the number of components extracted by CLEANUP.

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA
PAGE NUMBER
EPA-287 (Cin.)
(9-78)

background because of insufficient peak area or by not having two rising and falling points. The same problems occur when several peaks overlap that have ions in common. Because the MS sees only the sum of the ions in common, no distinct peaks are seen for these components. Also, when local background is subtracted, these ions may be lost or their areas severely miscalculated.

A second problem is that as the end of the GC temperature program is reached, the widths of the eluted peaks become greater. As these widths increase, a point is reached at which the peaks no longer have acceptable slope ratios and are eliminated. Finally, noisy data peaks may be eliminated due to noise spikes on either the ascending or descending side of the peak.

## SPECTRUM MATCHING

The spectrum matching portion of the data analysis system has undergone the least modification. PBM has been evaluated in the literature[5] and has been used in the Athens Laboratory for several years. Before the start of this project, PBM was operational on an in-house PDP 11/70.

Selection of a data base of reference spectra for use with PBM involved no small problem. Three data bases were available: the Wiley collection, the National Bureau of Standards collection, and the EPA master collection. The Wiley library contains 30,476 spectra of 30,476 compounds; the NBS library 25,025 spectra of 25,025 compounds; and the EPA master list contained $\sim$ 40,000 spectra of $\sim$ 32,000 compounds and is the master list of spectra from which the NBS library was taken.

Because the GC/MS data used in this study came from a great variety of sources it was thought that "duplicate" spectra, _i.e._ multiple spectra of the same compound that differ slightly due to run conditions, in the data base would be of some help in the matching process.

Table II compares the matching ability of two data bases on the same spectra. As can be seen, Data Base II (the EPA master data base) enjoys a distinct advantage over Data Base I (the NBS library) for the cases mentioned in the table. Comparison of the matches suggested by the two data bases with the manual identification shows the superior ability of Data Base II. Data generated using the Wiley library showed similar shortcomings to the NBS library.

In cases where the identical matched spectrum occurred in all the libraries, as was generally the case, no problems were observed. In some cases when spectrum extraction or run condi-

TABLE II. COMPARISON OF PBM MATCHING ABILITY FOR TWO DATABASES. DATABASE I IS NBS LIBRARY, DATABASE II IS EPA MASTER DATABASE.

| Database I (K value, missing masses) | Database II (K value, missing masses) | Manual ID |
|---|---|---|
| Toluene (75+) | Toluene (75+) | Toluene |
| 7-oxabicyclo 2,2,1 heptane (49+) | 2-cyclohexene-1-ol (76+) | 2-cyclohexene-1-ol |
| Phthalide (56, -2) | Methyl benzoate (69+) | Methyl benzoate |
| Hexacosanoic acid (102, -3) | Octadecanoic acid (105+) | Octadecanoic acid |

(+ means that the molecular ion was matched within the proper intensity tolerance) tions slightly affected the spectrum, only if there were duplicate spectra did a match occur. Thus, the EPA master library was chosen for use in this work.

The reverse search approach of PBM has also proven useful when analyzing environmental samples with an automated system. Figure 8 is an example of a mixed spectrum obtained by CLEANUP from a VOA standard run. Although the two compounds (cis-1,3-dichloropropene and 1,1,2-trichloroethane) are not resolved, PBM was able to match both compounds in this spectrum. This ability of a reverse search to recognize components of mixed spectra is clearly an advantage.

## RESULT COLLATOR

The combination of MS data with retention indices provides a powerful tool in the automated extraction and identification of components in GC/MS data. The use of retention indices coupled with MS data for identification purposes was reported by Nau and Biemann[8] and Sweeley et al.[9] Subsequently, automated systems in several laboratories[10-12] have been reported that rely on retention data as well as MS data for computer aided identificaton of a known set of organic compounds. Table III shows RRT variation versus K variation for several compounds. Although the match quality parameters vary from 45 to 100, the RRT variation is only 0.01 to 0.03. RRTs are particularly important in the identification of compounds such as alkanes or alcohols, which have little or no molecular ion and exhibit highly similar spectra. This is shown in Figure 9, which compares spectra and RRTs of a $C_{13}$ and a $C_{14}$ alkane.

TABLE III. COMPARISON FOR RRT VS. K VARIATION FOR SELECTED MATCHED

| Compound | Range of RRT | Range of K |
|---|---|---|
| Dioctylphthalate | 0.03 | 45 - 100 |
| Phthalide | 0.01 | 57 - 77 |
| Toluic Acid | 0.03 | 48 - 85 |

In this project, RRTs have been such a sensitive indicator that we have been able to distinguish contractor laboratories that have not followed the prescribed temperature programming conditions. Table IV shows a comparison of mean RRTs for two laboratories with the mean RRTs observed for all other laboratories for the same compounds. Based on these discrepancies,

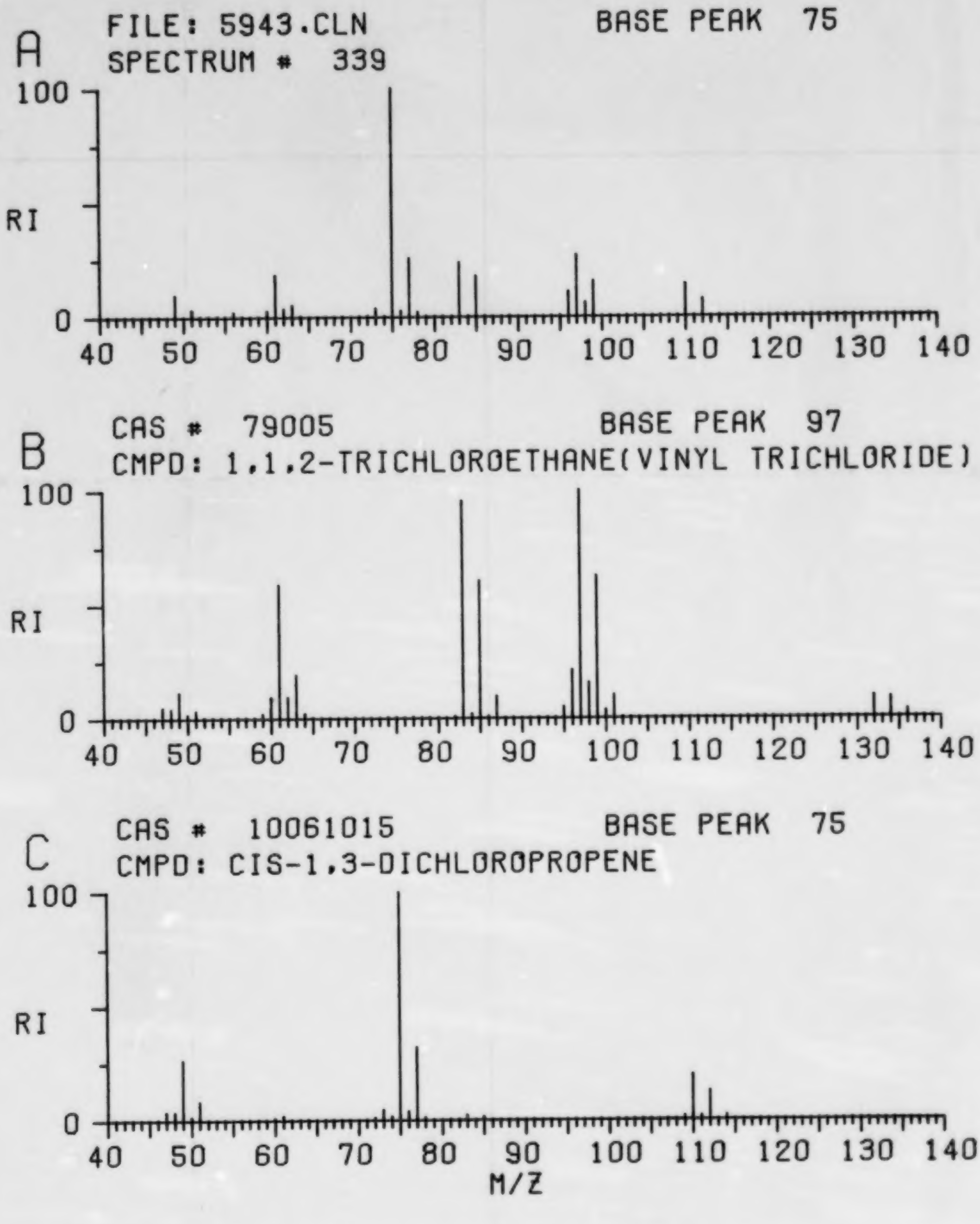


Figure 8. A. Mixed spectrum extracted by CLEANUP from a VOA standard run. B & C. Library spectra indicated as matched by PBM.

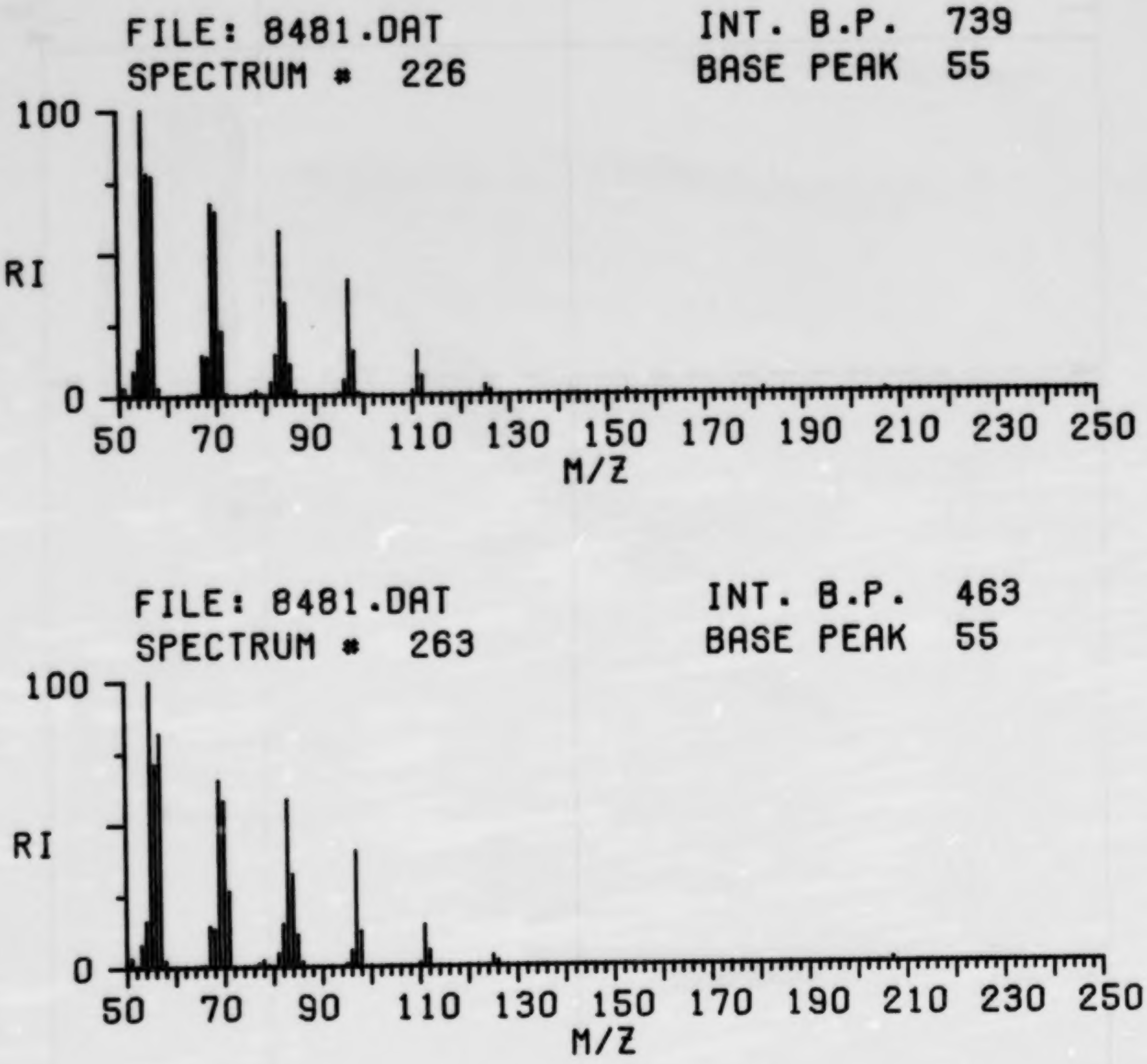


Figure 9. Comparison of spectra for two alkanes. Top spectrum (#226) is tridecane (RRT=.510). Bottom spectrum (#263) is tetradecane (RRT=.591).

it was discovered that these laboratories were indeed not temperature programming in the same manner as the others. As noted in the table, a linear correction factor is applied to the data from these laboratories to facilitate automatic spectra matching by HISLIB.

TABLE IV. MEAN RRTs FOR TWO LABORATORIES BEFORE AND AFTER CORRECTION, COMPARED WITH MEAN RRTs FOR OTHER LABORATORIES FOR B/N's RUN ON 1% SP-2250.

| Compound | RRT before Correction | RRT after Correction[1] | RRT for all other labs |
|---|---|---|---|
| Xylenes | .041 | .137 | .136 |
| 2-n-Butoxyethanol | .090 | .180 | .166 |
| $C_3$ Benzenes | .111 | .200 | .208 |
| Naphthalene | .375[2] | .438 | .457 |
| Methyl Naphthalenes | .491 | .542 | .553 |
| $C_2$ Naphthalenes | .612 | .651 | .663 |
| Fluorene | .796[2] | .816 | .838 |

[1] corrected using following formula:

$$RRT_{corrected} = RRT_{orig} + 0.1\ (1 - RRT_{orig}) \quad (\text{for } RRT_{orig} < 1)$$

[2] one observation only

Before actual runs were processed through the system, the historical library was initialized with data from standard runs. Several acid, base/neutral and VOA priority pollutant standard runs from the various contract laboratories were processed and the compounds entered in the historical library after being visually inspected and checked against the protocol for RRT consistency. Whenever possible, mean RRT values were calculated. All entries from standard runs were flagged so they would not be counted in the frequency of occurrence statistics.

As discussed above, several compound classes are difficult to distinguish without RRT data. Therefore, standards were obtained for the homologous series compound classes listed in Table V. These standards were run on the various GC columns and the compounds entered to the historical library.

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 287)
TYPING GUIDE SHEET

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA

TABLE V. HOMOLOGOUS SERIES STANDARDS

$C_6$ - $C_{19}$ n-Alkanes
$C_{19}$ - $C_{40}$ n-Alkanes
$C_6$ - $C_{10}$ Alkenes
$C_8$ - $C_{22}$ Alkenes
$C_4$ - $C_{22}$ n-Alcohols
$C_3$ - $C_{16}$ Aldehydes
$C_4$ - $C_{14}$ Primary Amines
$C_4$ - $C_{18}$ Secondary and Tertiary Amines
Benzenoid Hydrocarbons
Dicarboxylic Acids
Dimethyl Esters of Dicarboxylic Acids
$C_3$ - $C_{18}$ Fatty Acids
$C_3$ - $C_{12}$ Glycols
$C_3$ - $C_{10}$ Glycol Ethers
Low Boiling Esters
$C_3$ - $C_{19}$ Methyl Ketones
Phenols
Phthalate Esters

These standard runs, then, furnished a starting point for construction of the library. Compounds have been added to the library over the course of the project with 1565 unique compounds tentatively identified to date. The identification rate (the percent of spectra extracted by CLEANUP that have been tentatively identified) has increased as a function of the number of compounds in the library, and has plateaued at 32%. This is demonstrated in Table VI, which compares project status as of Dec. 31, 1979 with the results to date. At present, over 232,000 spectra have been extracted by CLEANUP, of which approximately 74,500 have been tentatively identified.

TABLE VI. PROJECT STATUS

| | | As Of | |
|---|---|---|---|
| | 12/31/79 | 07/18/80 | 10/01/81 |
| Unique Compounds in HISLIB | 620 | 945 | 1565 |
| Spectra Tentatively Identified | 8117 | 22978 | 74508 |
| Spectra Extracted by CLEANUP | 45574 | 99894 | 232366 |
| Identification Efficiency | 17.8% | 23.0% | 32.1% |

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 227)
TYPING GUIDE SHEET

A comparison of the most frequently occurring compounds as of Dec. 31, 1979, in Table VII with the lists in Table VIII and IX shows little compound variation with some shift in the order of frequencies. Duplicate occurrences of compounds in the acid, B/N and VOA fractions of the same sample have been eliminated prior to generation of these lists. The extracting solvent, dichloromethane, is never counted in acid or B/N runs. Background contaminants found in blank runs, however, have not been subtracted from these statistics.

TABLE VII. MOST FREQUENTLY FOUND COMPOUNDS
(AS OF 12/31/79)

| HISLIB Frequency | Compound Name |
|---|---|
| 266 | 1-(2-Butoxyethoxy) Ethanol |
| 180 | Cresols |
| 168 | *Phenol |
| 152 | Toluene |
| 145 | *Tetrachloroethylene |
| 134 | *Benzene |
| 120 | Benzoic Acid |
| 113 | *Dichloromethane |
| 102 | *Naphthalene |
| 92 | p-Xylene |
| 91 | *Ethyl Benzene |
| 88 | 1-Methyl Naphthalene |
| 85 | Octadecane |
| 81 | *Di-n-Octyl-Phthalate |
| 80 | Hexadecane |
| 75 | 2-Methyl Naphthalene |
| 75 | 2-Butoxyethanol |
| 75 | 1-Methyl-2-Ethyl Benzene |

* Priority Pollutant

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 287)

TYPING GUIDE SHEET

## TABLE VIII. 25 MOST FREQUENTLY FOUND COMPOUNDS (AS OF 7/18/80)

| HISLIB Frequency | Compound Name |
|---|---|
| 489 | 2-(2-Butoxyethoxy) Ethanol |
| 485 | *Tetrachloroethylene |
| 436 | Cresols |
| 414 | Xylenes |
| 406 | Toluene |
| 349 | *Dichloromethane |
| 335 | *Phenol |
| 302 | *Benzene |
| 274 | Benzoic Acid |
| 263 | *Ethyl Benzene |
| 262 | Octadecane |
| 256 | Hexadecane |
| 253 | *Trichloroethylene |
| 230 | *Naphthalene |
| 214 | 1-Methyl Naphthalene |
| 210 | *Butyl Phthalate |
| 206 | Pentadecane |
| 194 | *Bis(2-Ethylhexyl) Phthalate |
| 188 | *Di-n-Octyl-Phthalate |
| 179 | *1,1,1-Trichloroethane |
| 175 | *2-Methylnaphthalene |
| 174 | *Chloroform |
| 151 | *Diethyl o-Phthalate |
| 151 | 2-Butoxyethanol (Butyl Cellosolve) |
| 145 | Alpha-Terpineol |

* Indicates Priority Pollutant.

The increase in the number of unique library compounds over the last 7 months is largely due to the processing of runs from a greater variety of industrial categories. Profiles of samples analyzed by industrial categories as of Dec. 31, 1979 and July 18, 1980 are shown in Tables VIII and IX, respectively. Comparison of these tables shows that while POTW is still the most frequently analyzed industry, other categories such as timber products, organics and plastics, and nonferrous metals have increased considerably. Examples of a few "industry specific" compounds are listed in Table X, which shows the total number of HISLIB entries and the percentage that have been tentatively identified in a given industry.

The MISS library collates results for unidentified spectra that appear to reoccur within narrow RRT windows (± 0.1 RRT

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 287)
TYPING GUIDE SHEET

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA

## TABLE IX. MOST FREQUENTLY FOUND COMPOUNDS (AS OF 10/01/81)

DROPPED HEAD; BEGIN SECTIONS HERE

| HISLIB Frequency | Compound Name |
|---|---|
| 1497 | *tetrachloroethylene |
| 1442 | *toluene |
| 1439 | *dichloromethane |
| 1262 | 2-(2-butoxyethoxy) ethanol |
| 1147 | *bis (2-ethylhexyl) phthalate |
| 1103 | *chloroform |
| 958 | xylenes |
| 914 | *benzene |
| 848 | cresols |
| 827 | *di-n-butylphthalate |
| 763 | methyl naphthalenes |
| 758 | *phenol |
| 757 | *di-n-octylphthalate |
| 681 | *ethylbenzene |
| 664 | *trichloroethylene |
| 631 | *1,1,-trichloroethane |
| 627 | n-octadecane |
| 617 | n-hexadecane |
| 528 | benzoic acid |
| 515 | *naphthalene |
| 514 | *diethylphthalate |
| 473 | α-terpineol |
| 446 | n-pentadecane |
| 401 | n-hexane |
| 395 | caffein |

BEGIN LAST LINE OF TEXT
3/8"
BOTTOM OF IMAGE AREA, OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS

EPA Form 2360-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 287)
TYPING GUIDE SHEET

## TABLE X. SELECTED HISLIB RESULTS INDICATING "INDUSTRY SPECIFIC" COMPOUNDS

| Compound | Total HISLIB Entries | % of Entries found in Industrial Category | Industry |
|---|---|---|---|
| Dimethyl Sulfone | 30 | 70% | Pulp & Paper |
| Caffeine | 395 | 91% | POTW |
| Furfural | 46 | 33% | Rum |
| Xanthene | 11 | 82% | Coal Mining |
| p-Toluenesulfonamide | 21 | 38% | Printing & Publishing |
| Triphenylphosphate | 44 | 20%<br>30%<br>23% | Nonferrous Metals<br>Foundries<br>Aluminum |
| Syringaldehyde | 41 | 71% | Pulp & Paper |
| Acetosyringone | 24 | 67%<br>25% | Pulp & Paper<br>Rum |
| Acridone | 11 | 55% | Timber Products |
| Dibenzofuran | 66 | 29%<br>23% | Coal Mining<br>Timber Products |
| Diphenylether | 116 | 34%<br>14%<br>25% | Organic Chemicals<br>Organics & Plastics<br>POTW |
| Veratryl Alcohol | 7 | 71% | Nonferrous Metals |

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 337)
TYPING GUIDE SHEET

unit). To date, almost 159,000 such unidentified spectra have been entered into the MISS library. There are 606 unidentified components that have been found with a frequency of 15 or greater. These represent 10% of the total spectra cataloged in the MISS library. A sample entry is shown in Figure 10. This unknown has occurred 11 times in acid extracts run on 1% SP-1240 DA and is thought to be $D_8$ anthraquinone from degradation of the internal standard $D_{10}$ anthracene. This spectrum has been observed occasionally when manually locating internal standards, and because no deuterated compounds exist in the reference library, has been passed to the MISS library where recurrences have been collected.

## SUMMARY

Automated survey techniques for processing GC/MS data appear to work well on most of the data encountered in this work. Whereas the sensitivity of the CLEANUP-PBM package is not as great as that of a reverse search for specific ions, it is sufficient for the tentative identification of the major components in a sample. CLEANUP-PBM is cost effective when compared to an operator's finding peaks, subtracting background, and matching the spectrum manually or using a computer search.

Use of a historical library to catalogue data collected over long periods of time aids identification immensely by adding the dimension of GC retention data. Confidence in a tentative identification is heightened if there is corroborating GC retention data. This combination of spectral and retention data can be used effectively to catalogue and highlight recurring unidentified substances for future study.

Two areas are the subject of continuing study. The proper compensation of background by CLEANUP is of concern since errors in intensity calculations reduce the chance that PBM will find a match for the spectrum. Studies underway indicate that background compensation similar to that used in PEAK would be effective in CLEANUP. Implementing such a routine is under study.

A second area of concern is the various parameters that must be set for CLEANUP. These parameters that pertain to peak shape and minimum area are inflexible during a given data run. Thus, changing chromatographic conditions that affect peak size and shape may cause CLEANUP to miss good data. The automatic setting of these parameters during data processing is desirable and is under study at present.

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN EPA FORM 187)
TYPING GUIDE SHEET

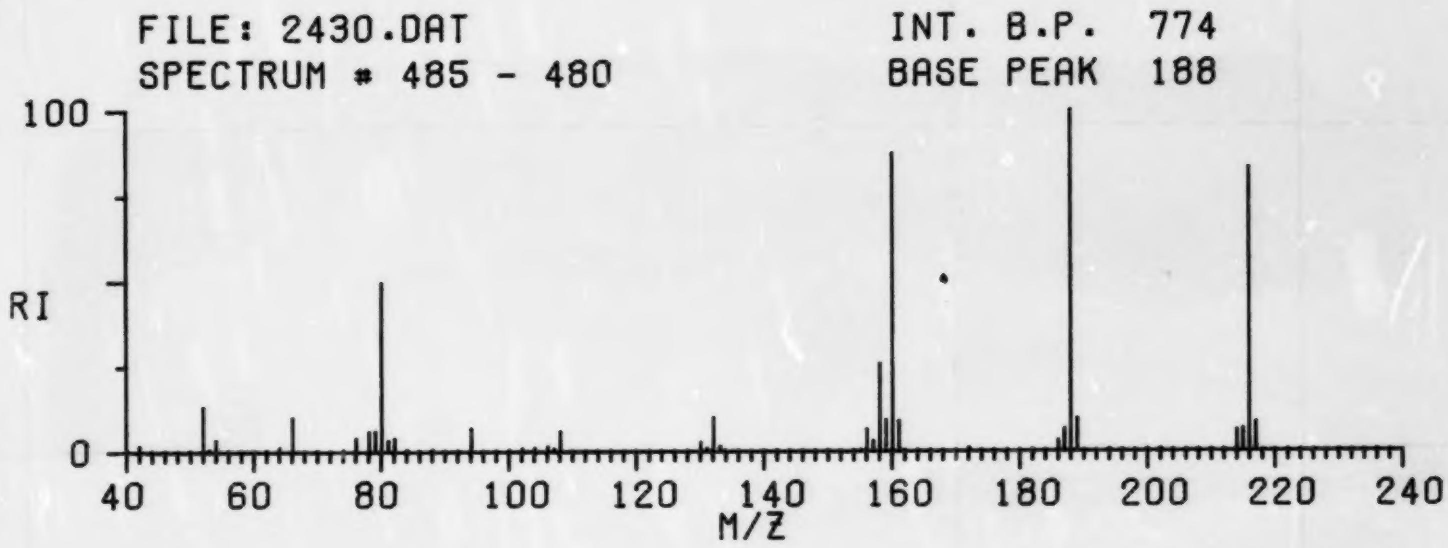


Figure 10. Sample spectrum from MISS library thought to be $D_8$ anthraquinone from degradation of internal standard $D_{10}$ anthracene. This unknown has occurred 11 times in acid extracts run of 1% SP-1240 DA.

## REFERENCES

1. Shackelford, W. M. and L. H. Keith. Evolution of the Priority Pollutant List from the Consent Decree. In: Proceedings of the Second Open Forum on Management of Petroleum Wastewater. EPA-600/2-78-058, U.S. Environmental Protection Agency, Ada, Oklahoma, 1978. pp. 103-109.

2. Keith, L. H. and W. A. Telliard. Priority Pollutants I - A Perspective View. Environ. Sci. Technol. 13(4):416-423, 1979.

3. Sauter, A. D., L. F. Faas, G. R. Kurth, W. M. Shackelford, and D. M. Cline. Applications of a Data Adaptive Background Subtraction Technique to the GC/MS Analysis of Waste Water Samples. American Chemical Society National Meeting, San Francisco, California, August 27, 1980.

4. Dromey, R. G., Mark J. Stefik, Thomas C. Rindfleisch, and Alan M. Duffield. Extraction of Mass Spectra Free of Background and Neighboring Component Contributions from Gas Chromatography/Mass Spectrometry Data. Anal. Chem. 48(9):1368-1375, 1976.

5. Pesyna, Gail M., Rengachari Venkataraghavan, Henry E. Dayringer, and F. W. McLafferty. Probability Based Matching System Using a Large Collection of Reference Mass Spectra. Anal. Chem., 48(9):1362-1368, 1976, EPA Resarch Report No. EPA-600/4-76-046, October 1976.

6. Smith, Dennis H., Michael Achenbach, William J. Yeager, Patricia J. Anderson, William L. Fitch, and Thomas C. Rindfleisch. Quantitative Comparison of Combined Gas Chromatography/Mass Spectrometric Profiles of Complex Mixtures. Anal. Chem., 49(11):1623-1632, 1977.

7. Pesyna, G. M., F. W. McLafferty, Rengachari Venkataraghavan, and H. E. Dayringer. Statistical Occurrence of Mass and Abundance Values in Mass Spectra. Anal. Chem., 47(7):1161-1164, 1975.

8. Nau, H. and K. Biemann, Computer-Assisted Assignments of Retention Indices in Gas Chromatography-Mass Spectrometry and its Application to Mixtures of Biological Origin, Analytical Chemistry, 46(3):426-434, 1974.

9. Sweeley, C. C., N. D. Young, J. F. Holland, and S. C. Gates, Rapid Computerized Identification of Compounds in Complex Biological Mixtures by Gas Chromatography-Mass Spectrometry, Journal of Chromatography, 99:507-517, 1974.

10. Smith, Dennis H., Michael Achenbach, William J. Yaeger, Patricia J. Anderson, William L. Fitch, and Thomas C. Rindfleisch. Quantitative Comparison of Combined Gas Chromatography/Mass Spectrometric Profiles of Complex Mixtures. Anal. Chem., 49(11):1623-1632, 1977.

11. Tsuchiya, Yoshio, J. Boulanger, and K. Sumi. Identification of Chemical Compounds by Computer Matching of Low Resolution Mass Spectra and Retention Indices, Chromatographia, 10(3):154-156, 1977.

12. Blaisdell, B. Edwin. Automatic Computer Construction, Maintenance, and Use of Specialized Joint Libraries of Mass Spectra and Retention Indices from Gas Chromatography-Mass Spectrometry Systems, Analytical Chemistry, 49(1):180-186, 1977.

# APPENDIX I.

# GC/MS DATA SURVEY SYSTEM
# USER'S GUIDE
# (VERSION III)

by

F. O. Burchfield, G. R. Kurth,
and L. A. Faas
Computer Sciences Corp.
Applied Technology Division
6565 Arlington Blvd.
Falls Church, VA  22046

# TABLE OF CONTENTS

**Title** **Page**

Introduction 38

Data Handling Procedure 43
  Receive Tapes (1) 43
  Log Tapes and Verification (2) 43
  General Description of Tape Conversion Program (3) 50

INCOS Tape Format 51

System 150 Format 53

Hewlett Packard Format 54

Tape Conversion Errors 55

User Documentation for Tape Conversion Programs 56
  Create Control File (3A) 56
  Setting up and Running Tape Conversion Programs (3B) 58
  Plotting RGC Plot Files for Runs (3C) 59
  Updating the INFORM Database (3D) 60

General Description of Control Program (4) 61
  Create Control File (4A) 65
  Create and Submit Batch File (4B and 4C) 66

Computer Analysis Modules (5) 69
  RETRIEVE 69
  BUILD 69
  STANDARD 69
  PEAK 70
  MODELS 70
  CLEANUP 70
  PBM 70
  TIMSEK 71
  HISLIB 71

Historical Library Algorithm 73
  Pass 1 73
  Pass 2 75

Graphics (6) 83

Report Generation (7) 88

References 93

Appendix A

- INFORM Data Base Definitions 94
- INFORM Data Base Structure - TAPE.LOG 97

Appendix B

- Description of SPECIF.DAT 98
- Description of Assignment for Logical Unit Numbers for Tape Conversion 99
- Documentation on FORTRAN Callable QIO 100
- Internal Tape Format 102
- Description of Plot Batch File to Plot RGC's 104
- Description of AERL Update Command File 105
- INCOS Tape format 106
- Hewlett Packard Tape Format 108

Appendix C

- File Structures 109
  - EXPLIC.TIM 109
  - EXPLIC.PBM 110
  - EXPLIC.CLN 111
  - EXPLIC.SEK 112
  - HISLIB.RRT 114
  - HISLIB.HIT 115
  - EXPLIC.HLD 117
  - EXPLIC.BCK 117
  - GCCOL.DAT 118
  - ANALAB.DAT 118
  - INTSTD.DAT 118
  - RUNNAME.STD 119
  - MASS.TMP 119
  - MASS.STD 120
  - GCWINDOW.MIS 121
  - RRT.MIS 122
  - HISLIB.MIS 123
- LIST.MAS File Format 124
  - Header Record Format 124
  - Data Record Format 126
  - Finnigan File Formats 127
  - RUNNAME.PAR File Format 127
  - RUNNAME.DIR File Format 131
  - RUNNAME.DAT File Format 131
- Description of Files on Output Tape 132
- Description of FOR005.DAT Created by BATCHJOB Used During Run Processing 134

Appendix D

Computer Analysis Modules - Inputs and Outputs

- RETRIEVE 135
- BUILD 136
- STANDARD 137
- PEAK 138
- MODELS 139
- CLEANUP 140

PBMUNK 141
PBM 142
TIMSEK 143
HISLIB 144
SPECTRA 145
GCPLOT 146

APPENDIX E

ADDITION OF SPECTRUM TO THE PBM Database 147
SPEC.DAT File Structure 148
Records for Mass/Int Pairs of Spectrum 149

## INTRODUCTION

The GC/MS data survey and analysis system was assembled by Computer Sciences Corporation (CSC) for the Athens Environmental Research Laboratory. GC/MS data collected by analytical laboratories is transferred to the system via magnetic tape. As soon as the tapes are received, the associated descriptive data are logged into a database management system, INFORM, and verified to ensure the validity of the data. Various analytical laboratories use different data collection systems that are incompatible with each other. Therefore, tapes are reduced to a standard format that uses the magnetic tapes more efficiently and creates a reconstructed gas chromatogram for each run.

The computer analysis consists of four major sections--CLEANUP, PBM, TIMSEK, and HISLIB. CLEANUP extracts spectra for each GC/MS run. These spectra are input to PBM for matching against the spectrum library. Relative retention time(s) and relative concentration(s) are then computed by the program TIMSEK. The output from TIMSEK is then processed by HISLIB(1), which checks the compounds suggested by PBM and their relative retention times from TIMSEK against the historical library of compounds for matches or tentative identifications. The historical library catalogs these tentative identifications and provides statistics on their frequency of occurrence. Also, a historical library of unidentified spectra is maintained.

CLEANUP (2) was originally written at Stanford University, Department of Chemistry and Genetics, and PBM (3) was written at Cornell University. CSC converted and fine-tuned these two programs for the inhouse EPA computer. All other modules were written by the CSC staff. The logging procedures utilize the INFORM data base management system.

Initially, all work for this project used a DEC PDP 11/70 with 256K words of memory and 2K words of cache memory, two 88 M-byte disk drives (DEC RP04), two magnetic tape drives (DEC TU16), a Calcomp 936 plotter, and a Tektronics 4014 graphics display terminal. All run processing functions were transferred to a PDP 11/34 multi-user dedicated system with 128K words of memory and 2K words of cache memory, two 88 M-byte disk drives (DEC RP04), and a Versatec 1200 printer/plotter. Both systems utilize the Interactive Applications System (IAS) operating system.

FIGURE I illustrates the procedure for handling the incoming data. The numbers in the figure identify the process steps that are described in the following sections.

All user responses in this manual are underlined.

(Version III of the GC/MS User's Guide replaces Version II dated November 1979.)

FIGURE I

CONSENT DECREE DATA FLOW

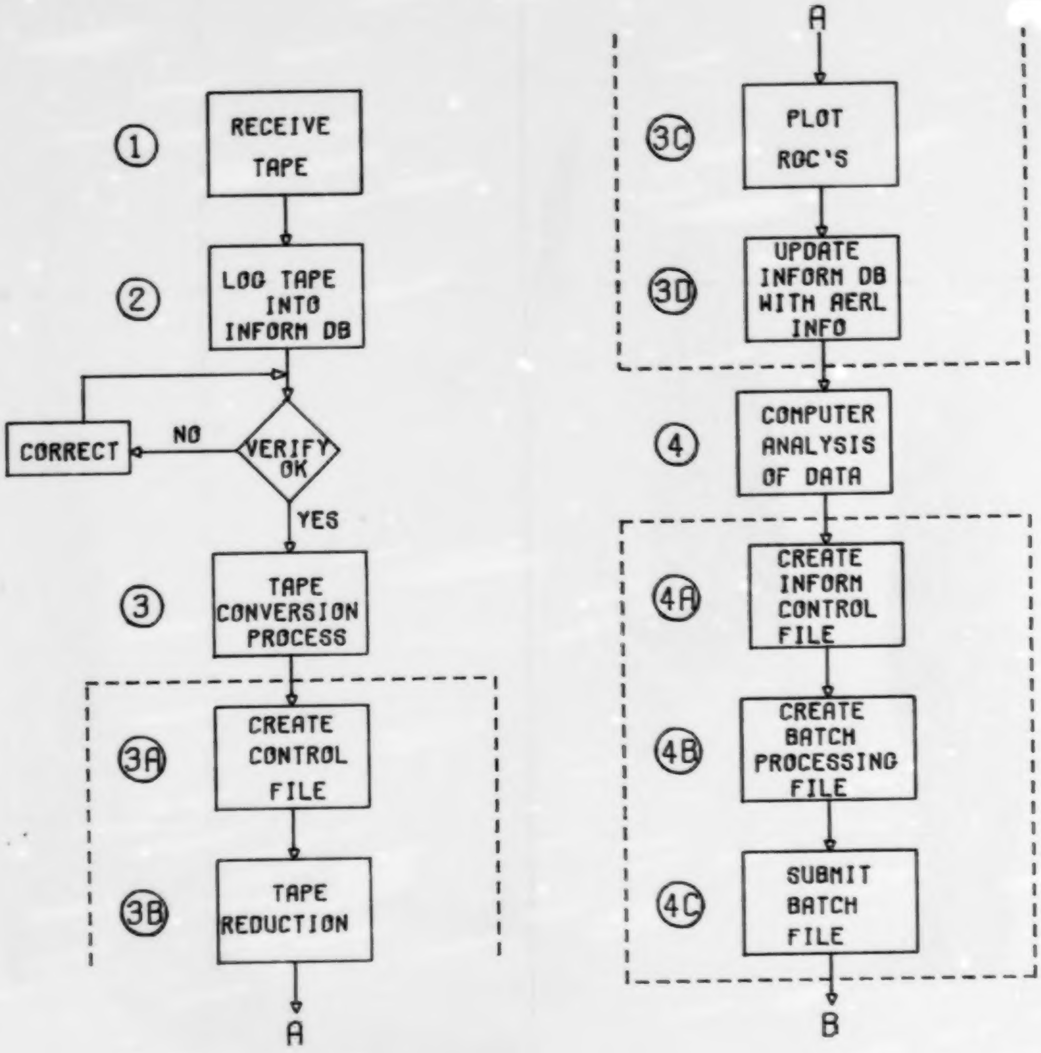
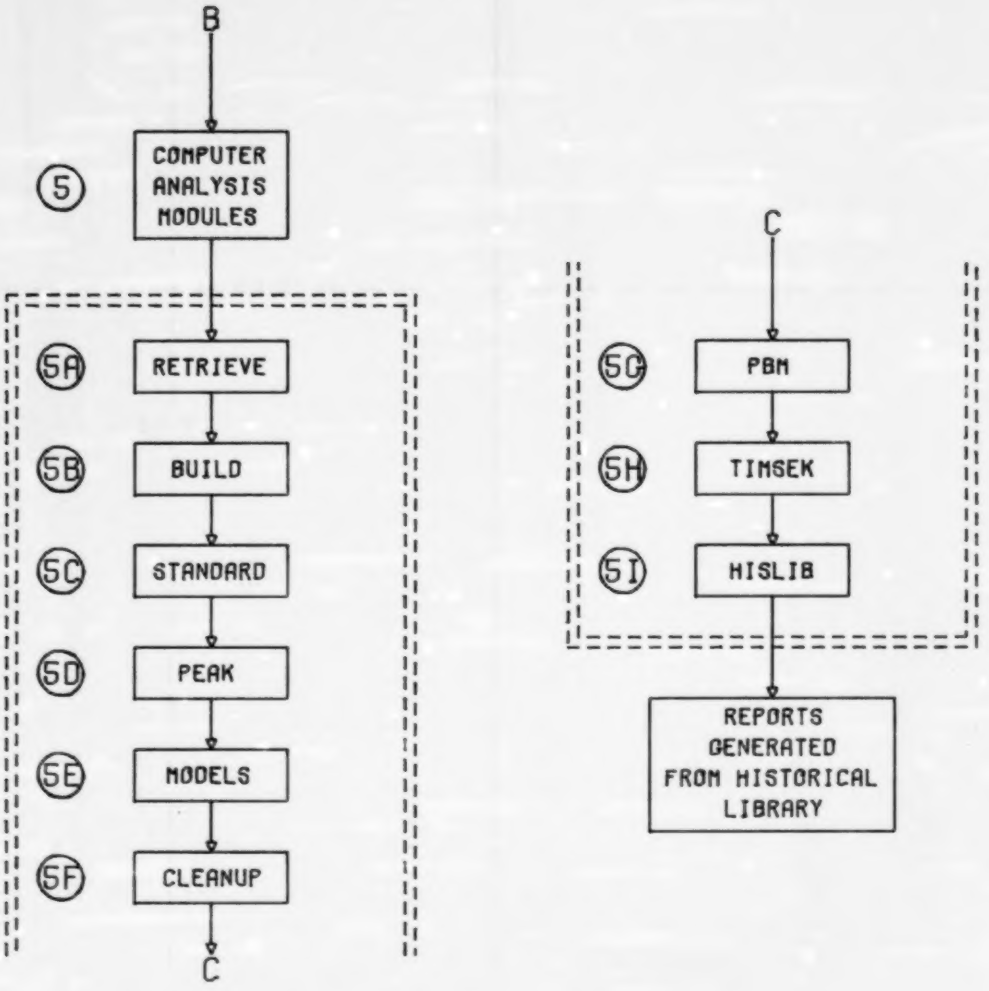

①
RECEIVE TAPE
②
LOG TAPE INTO INFORM DB
CORRECT
NO
VERIFY OK
YES
③
TAPE CONVERSION PROCESS
3A
CREATE CONTROL FILE
3B
TAPE REDUCTION
A
A
3C
PLOT ROC'S
3D
UPDATE INFORM DB WITH AERL INFO
④
COMPUTER ANALYSIS OF DATA
4A
CREATE INFORM CONTROL FILE
4B
CREATE BATCH PROCESSING FILE
4C
SUBMIT BATCH FILE
B

FIGURE I CONT.
CONSENT DECREE DATA FLOW


B
⑤
COMPUTER ANALYSIS MODULES
5A
RETRIEVE
5B
BUILD
5C
STANDARD
5D
PEAK
5E
MODELS
5F
CLEANUP
C
C
5G
PBM
5H
TIMSEK
5I
HISLIB
REPORTS GENERATED FROM HISTORICAL LIBRARY

## DATA HANDLING PROCEDURE

### Receive Tapes (1)

All of the magnetic tapes containing the GC/MS runs and their descriptive data sheets are mailed to the Project Officer of the data survey system at the Athens Laboratory. The tapes are entered in a custody log book and turned over to the CSC inventory control clerk.

### Log Tapes and Verification (2)

The first step of the tape logging process is to check the descriptive data sheets for completeness. In addition, the file headers for each run on the tape are checked for agreement with the data sheets and the tape is checked for parity, bad blocks, etc. If any discrepancies or problems are noted, the tape is returned to the Project Officer for corrections. If the sheets are in order, the tape is assigned an inhouse number. The associated GC/MS runs are then numbered for entry into the INFORM database, TAPE.LOG. The GC/MS descriptive data are now ready to be logged into the system via the following commands.

```
PDS> LOGIN LYN
PASSWORD?
PDS> INE DO LOG
```

A format for data entry will appear on the screen (FIGURE II). The descriptive data for one GC/MS run comprise a screen of information. Each run is assigned a number represented in the data base by RUN.NAME. (This is the reference number used throughout the computer analysis.) All the other information entered on the screen is obtained from the descriptive data sheets. (For field definitions of the data base, see Appendix A.)

A second program, TAPELOG, is available to log GC/MS descriptive data. This program inserts the internal standards and their concentrations, the GC column, GC, MS, and data system via a table lookup. Both programs log information to a temporary database, TAPE.LOG.TRANSACT, which is computer checked against itself and TAPE.LOG for possible duplicate runs before being updated into TAPE.LOG and converted.

The entered data are verified in one of two ways.

1) At a user terminal,

```
PDS> LOGIN LYN
PASSWORD?
```

```
PDS>  apout
ENTER BEGINNING RUN NAME?  <run name>
ENTER ENDING RUN NAME?     <run name>
```

The operator enters the appropriate information. The run listing will be printed on the line printer for the run names entered. These listings are a projection of the login screen format (Figure III).

FIGURE II

```
RUN NAME 1517     TAPE# ****    EXTRACT#____     EXTRACT LOCATION__

USER RUN NAME_________

                           USER FILE#_  USER SAMPLE#____________

EPA SAMPLE#______________  FRACTION# __  INDUSTRIAL CODE __

TREATMENT STAGE_____________________  DATE SAMPLED____

DATE RUN ____   SAMPLER CODE _       MS ______________

GC ______________               DATA SYSTEM ___________

GC COLUMN _____________   AERL TAPE# __     AERL FILE# __

CONC FACTOR__     INJECTION VOLUME_        ANAL LAB ____________

USER TAPE #_____________

INTERNAL STD CONC 1 _ 2 _ 3 _       INTERNAL STD 1 __________

INTERNAL STD 2 ________         INTERNAL STD 3 __________
```

## FIGURE III

```
RUN NAME    888    TAPE# 1663    EXTRACT# VOA      EXTRACT LOCATION VOA

USER RUN NAME 28010230

                          USER FILE# 11   USER SAMPLE# 28010230

EPA SAMPLE# 5-H-023                       FRACTION# VOA   INDUSTRIAL CODE 08

TREATMENT STAGE NOT KNOWN                              DATE SAMPLED 00/00/00

DATE RUN 06/20/78   SAMPLER CODE 20        MS FINNIGAN 3200

GC FINNIGAN 9500                DATA SYSTEM INCOS

GC COLUMN 0.2%CARBOWAX 1500 CARBOPAK C   AERL TAPE#           AERL FILE#

CONC FACTOR        INJECTION VOLUME          ANAL LAB ARLI

USER TAPE # ARLI #4

INTERNAL STD CONC 1 20 2 20 3          INTERNAL STD 1 BROMOCHLOROMETHANE

INTERNAL STD 2 1,4-DICHLOROBUTANE      INTERNAL STD 3
```

2) At a user terminal,

```
PDS>  LOGIN LYN
PASSWORD?
PDS>  INF LOAD TAPE_LOG_TRANSACT
PASSWORD?
@@@       (The    appropriate    password    must    be
entered)
**
```

The operator enters the appropriate INFORM commands to list the data for the run names displayed on the screen.

The second technique is for the more experienced INFORM user. (See INFORM manual for commands and Appendix A for field names.) After obtaining the listings, the operator verifies the data against the data sheets.

If corrections are necessary, the following commands are entered at a terminal:

```
PDS>  LOGIN LYN
PASSWORD?
PDS>  INF DO CORTRANS
RUN NAME?  <enter run name of information to be
corrected>
```

The operator enters the run name desired. All the information entered for that particular run will appear on the screen (in the original login screen format). Using the <return> key, the operator advances through the fields until the entry requiring correction is reached. The correction is made and the review is continued through the remaining fields.

ACCEPT, REJECT OR EDIT? Will appear on the screen.

The operator enters A, R OR E, whichever is appropriate, to return to the run name prompt. To exit, <return> is entered in response to RUN NAME?.

A second program, INF DO CORRECT, is available to correct data in TAPE.LOG.

## General Description of Tape Conversion Programs (3)

The GC/MS runs received from various sources are written to magnetic tape by different data collection systems --such as INCOS, System 150, and HP-- interfaced to the GC/MS instruments. These different data systems pose a problem in that their tapes are written in incompatible formats. To resolve this, the tapes are converted to a standard format and the validity of the information provided with the tapes is confirmed. These tapes provide a consistent format so that the major modules (CLEAN-UP, Plotting routines) can easily access the tapes.

In addition to creating new tapes in a standard format, these programs build a reconstructed gas chromatogram (RGC) plot file for each individual GC/MS run , a batch file for plotting the RGCs, and a batch file for updating the TAPE.LOG database with the tape number and file position of each converted run.

## INCOS Format

### INCOS Tape Format

Each 'INCOS tape' contains a variable number of GC/MS data files. The number of files depends on the length of each run and the length of the tape. The tape characteristics are:

```
800 BPI
ASCII
BLKSIZE=800
LRECL  =80
LABEL  =NL
9-TRACK
```

Each GC/MS run is preceded by a four-line header that provides information about the run and the conditions of the instrument and is followed by the number of scans taken during the run. Each scan is composed of one header line (80 characters) and one or more lines with up to 13 mass/intensity pairs per line. The last line has a 00 mass/intensity terminator. The end of the run is signaled by an end-of-file mark on the tape. For a more detailed description of the file structure, see Appendix B.

A general description of the program flow is as follows: The main program initially opens the input control file (SPECIF.DAT), batch plotting file (PLOT.BIS), INFORM update file (AERL.DAT), and a temporary storage file for intermediate results. Maximum values are located for the output file header. The next step is a machine-dependent routine that assigns the tape drives to logical units, rewinds the tapes, and checks that the drive and tape characteristics are correct. Each run is then processed as an independent step, and in the event that an error occurs during the processing of a run, the program will abort that run and continue with the next. This program flow can be interrupted by the following fatal errors:

1) End of tape reached on output tape before program execution has been completed.
2) End of tape reached on input before completing all runs specified by the control file.
3) Error when output tape was written, i.e. parity error.

When the first run is being processed, the output tape is positioned after the last file written to tape, otherwise there is no movement of the output tape to position it with each new run.

NOTE: For more detailed system documentation see program listings. Description of input and output file structures can be found in Appendix B.

## System 150 Format

As with the INCOS tapes, each System 150 tape contains a variable number of files depending on the length of the tape and the length of each individual run file. The System 150 tape structure differs from the INCOS tapes in that the PDP-8 magtape controller processed the files in 'core dump mode'. Each 12-bit PDP-8 word is divided into two 6-bit bytes that are transferred to magnetic tape by the controller. The two high order bits of each byte written to tape are set to zeros, and the remaining 6-bits are used to contain one-half of the PDP-8 12-bit word. This requires that the two high-order zero bits be removed from the low byte when processing a numeric value. This process is only required for the numeric values and not for characters because each character per byte can be represented with 6-bit ASCII code.

The general program flow for the System 150 tapes follows that of the INCOS tape system. A description of the input and output file structures can be found in Appendix B and a more detailed system documentation may be found in the program listings.

## HEWLETT PACKARD (HP) FORMAT

### Hewlett Packard Format

As with the INCOS tapes, each Hewlett Packard tape contains a variable number of files depending of the length of the tape and each individual run file. The HP tape is recorded at 800 BPI in ASCII format, and has a logical record size of 80 bytes and a block size of 4000 bytes. See Appendix B for detailed file structure.

The general program flow for the HP tapes follows that of the INCOS tape system. A description of the input and output file structure can be found in Appendix B and a more detailed system documentation may be found in the program listings.

# ERROR CODES

## Tape Conversion Errors

1) ERROR DURING OUTPUT TO TAPE
Occurrence of this error causes termination of program execution. All runs completed before this error are retrievable from output tape. Present run and all subsequent runs have not been processed and must be reprocessed.

2) END OF VOLUME ENCOUNTERED DURING OUTPUT TO TAPE
Occurrence of this error causes termination of program execution. Present run is deleted from output tape and must be rerun along with all subsequent runs using a new output tape.

3) END-OF-VOLUME ENCOUNTERED ON INPUT
Occurrence of this error causes termination of program execution. Current run has not been completed and is not accessible by this program. In other words, the run is lost.

4) BAD BLOCK
Processing of present run is aborted and is not transferred to output tape. Execution continues with next run.

5) INSUFFICIENT MEMORY SPACE FOR SPECTRUM
A spectrum in the run has more mass/inten. pairs than space available for storage in memory (750 pairs). Conversion of run continues but spectrum in error is truncated (System 150).

6) INVALID BLOCK LENGTH
Block size is not a multiple of the logical record length. Processing of present run is aborted and is not transferred to output tape. Execution continues with next run.

7) RUN NAME OUT OF SEQUENCE
Tape has been spaced forward past the current run to be processed. This run is not processed and execution continues with next run (System 150).

8) INSUFFICIENT MEMORY SPACE FOR RGC PLOT
The present run has more scans than space available for storage in memory (1500 scans). Conversion of run continues but RGC plot is incomplete.

## USER DOCUMENTATION FOR TAPE CONVERSION PROGRAMS

Create Control File (3A)

In Building the Control File, the following commands are used:

```
PDS> LOGIN LYN

PASSWORD?

PDS> DEL SPECIE.DAT;*

PDS> INE DO SPECIE


ENTER AERL TAPE NUMBER (4 DIGITS)?  <output tape number>


SYSTEM 150 TAPE (Y OR N)?  <IF SYSTEM 150 TAPE>

ENTER TYPE OF MAG TAPE CONTROLLER (TC58 OR WANGCO)?  <Provided by
Chemist>

ENTER VOA TIME OFFSET (SECONDS) FOR THIS TAPE?  <Provided by
Chemist>

ENTER B/N TIME OFFSET (SECONDS) FOR THIS TAPE?  <Provided by
Chemist>

ENTER ACID TIME OFFSET (SECONDS) FOR THIS TAPE?  <Provided by
Chemist>

ENTER PESTICIDE TIME OFFSET (SECONDS) FOR THIS TAPE?  <Provided
by Chemist>

ENTER DAI (DIRECT AQUEOUS INJECTION) TIME OFFSET (SECONDS)
FOR THIS TAPE?  <Provided by Chemist>

VOA OFFSET - <VALUE>
PEST OFFSET - <VALUE>
DAI OFFSET - <VALUE>
ACID OFFSET - <VALUE>
B/N OFFSET - <VALUE>

ARE THESE CORRECT?  <Enter Y or N>
```

(If INCOS tape, control proceeds to this point immediately)

```
OPTIONS (1)  ENTER TAPE NUMBER

        (2)  ENTER RUN NAMES

        (3)  EXIT

(1)  ENTER OPTION NUMBER?  1

     ENTER TAPE NUMBER?  <Enter appropriate tape number
     TAPE_NO>

(2)  ENTER OPTION NUMBER?  2

     ENTER RUN NAME?  <Start with first RUN_NAME>

     ENTER RUN NAME?  <Continue until all RUN_NAMEs have
     been given, carriage return exits>


(3)  ENTER OPTION?  3

     PDS> LOGO
```

## Setting Up and Running Tape Conversion Programs (3B)

The tape conversion program is performed by:

1) Mounting original data tape on Drive 0 with NO write ring

2) Mounting EPA internal format tape on Drive 1 with write ring in the tape

   The output tape may be in one of two states:

   a) New tape (If this is the case, then a job must be submitted with a new tape mounted on Drive 1 before any tape conversion may be run. The following commands need to be issued.

```
PDS> LOGIN LYN

PASSWORD?

PDS> SUB [203,377]NEW

PDS> LOGO
```

   b) Previously used tape with room available for more runs to be added.

3) Entering the following on a terminal that has not been logged on to an account.

```
PDS> @INCOS (for Incos Data Systems)

   or

PDS> @SYS150 (for System 150 Data Systems)

   or

PDS> @HPTAPE (for Hewlett Packard Data Systems)
```

The tape conversion produces the following:

a) Runs converted to 'internal' format on output tape
b) An RGC plot file for each run on DB1:[203,203]
c) INFORM control file for update of INFORM database
d) Trace back listings (DB1:LISTING.DAT)
e) Batch file on DB1:[203,203] for plotting plot files

## Plotting RGC Plot Files for Runs (3C)

The RGC's are plotted by the following procedure:

A. Copy all necessary RGC plot files to the account EGD ([203,203]) on device DB0: on the PDP 11/34 computer. The following procedure is used:

   1. Log into the account EGD as follows:

      ```
      PDS> LOGIN EGD
      ```

   2. MOUNT THE DRIVE AND COPY THE PLOT FILES TO DK0:

      ```
      PDS> MOUNT/NOOP/OVER:VOL DK0:  DISK

      PDS> COPY *.PLX DK0:*.*

      PDS> COPY PLOT.BIS DK0:*.*
      ```

   3. Dismount the drive.

      ```
      PDS> DISMOUNT DK0:
      ```

   4. Log into account EGD on the PDP 11/34.

      ```
      PDS> LOGIN EGD
      ```

   5. Mount drive and copy plot file to DB0:

      ```
      PDS> MOUNT/OVER:VOL DK0:  DISK

      PDS> COPY DK0:PLOT.BIS DB0:*.*

      PDS> COPY DK0:*.PLX DB0:*.*
      ```

B. Check the Versatec for a sufficient amount of paper and submit batch plotting file.

   ```
   PDS> DISMOUNT DK0:

   PDS> SUBMIT PLOT
   ```

## Updating the INFORM Database (3D)

The INFORM database is updated by the following commands:

```
PDS> LOGIN LYN

PASSWORD?

PDS> @AERL

PDS> LOGO
```

This step automatically updates the INFORM Tape Log Database and produces a trace back on the line printer.

## General Description of Control Program (4)

### To Run Computer Analysis of Data:

The computer data analysis program is used to build a batch control file for analyzing a large number of runs with minimal human intervention. This program is controlled by a file created by INFORM using the TAPE.LOG database and provides flags for each GC/MS module to be processed. A "0" flag indicates that a given module is to be executed. A "1" flag omits execution.

A standard computer analysis for a run consists of the following:

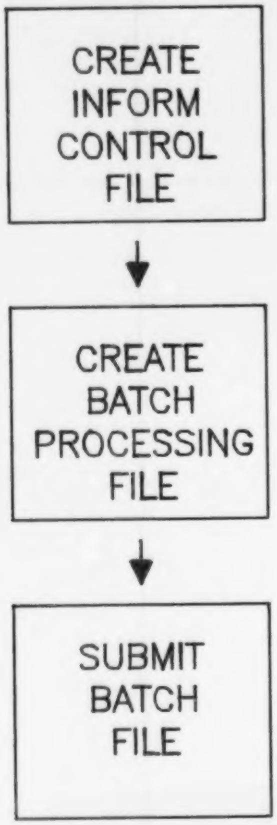
| FLAG | |
|---|---|
| 1 | Run RETRIEVE module |
| 2 | Run BUILD module |
| 3 | Run STANDARD module |
| 4 | Run PEAK module |
| 5 | Run MODELS module |
| 6 | Run CLEANUP module |
| 7 | Run PBM module |
| 8 | Run TIMSEK module |
| 9 | Run HISLIB module |
| 10 | Run GCPLOT module |
| 11 | Run SPECTRA module |
| 12 | Transfer and delete output files |

A complete computer analysis of a run consists of four major steps:

1) Retrieval of run from an standard format tape
   (Module - RETRIEVE)

2) Extraction of spectra for each run
   (Modules - BUILD, STANDARD, PEAK, MODELS, CLEAN)

## COMPUTER ANALYSIS OF DATA


CREATE INFORM CONTROL FILE
CREATE BATCH PROCESSING FILE
SUBMIT BATCH FILE

3) Matching of cleaned up spectra for each run against library spectra (Modules - PBMUNK & PBM)

4) Calculation of relative retention times and relative concentrations and updating of HISLIB databases. (Modules - TIMSEK and HISLIB)

The following represents the processing controlled by the batch file:

1) The file UNITS.DAT is created to assign logical unit numbers (LUN) for terminal input (TI:) and terminal output (TO:). All modules, except PBMUNK, are dependent on this file. The LUN's for these modules, when run in batch mode, are set to 5 for TI: and 6 for TO:. For interactive processing both are set to 5.

2) All output produced by the previous batch job is deleted. Only the output from CLEAN and HISLIB are printed following completion of the run. All output is retained on disk as a precautionary measure, e.g. lineprinter malfunction, until the next batch job is submitted.

3) The RETRIEVE module is executed for all GC/MS runs prior to the execution of the other modules.

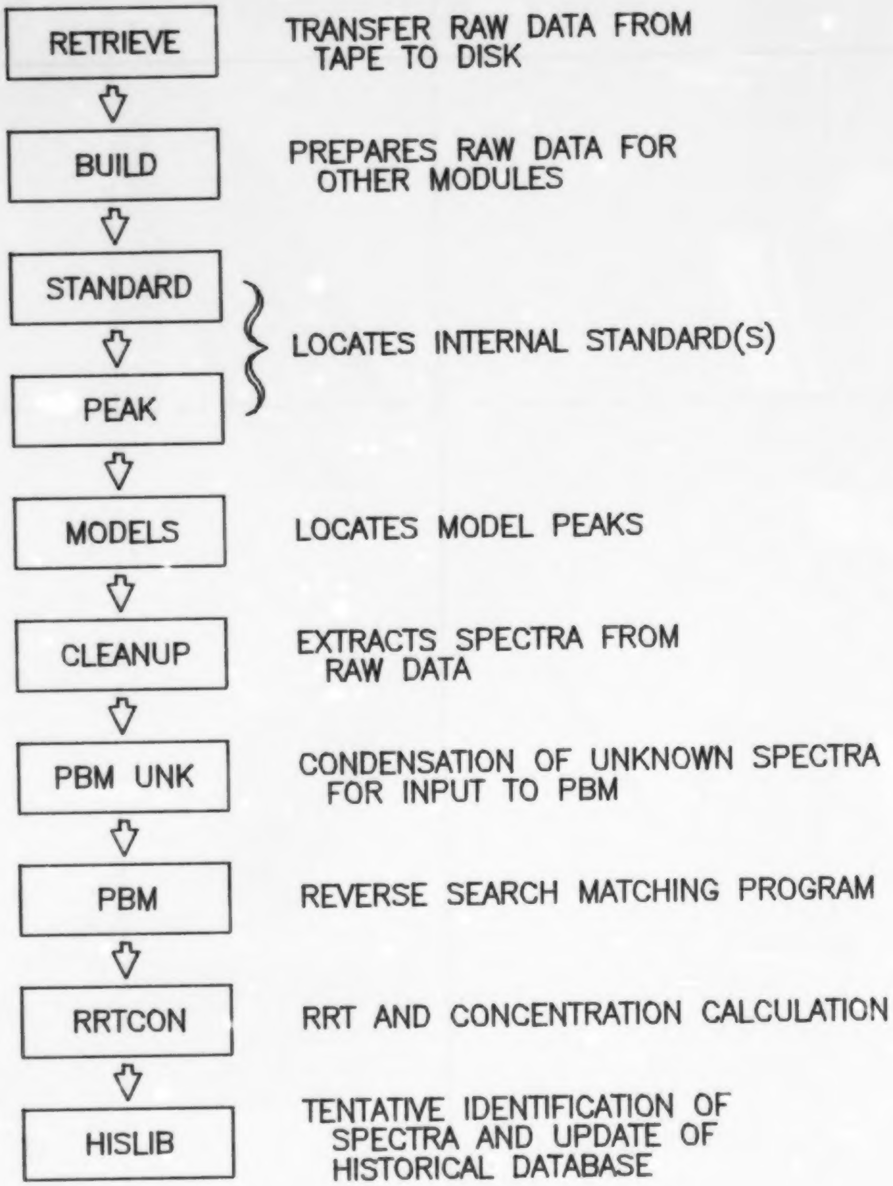
4) A file is created to simulate terminal input during the batch mode (FOR005 on the default disk).

5) Individual modules as specified by the run flags are then executed in order for each run.

If an error occurs in any of the modules then the processing of that run is terminated and program flow continues with the next run.

See pages 29-32 for instructions on running the modules interactively, and for a description of each module.

## COMPUTER ANALYSIS SYSTEM FLOW


RETRIEVE
TRANSFER RAW DATA FROM TAPE TO DISK
BUILD
PREPARES RAW DATA FOR OTHER MODULES
STANDARD
PEAK
LOCATES INTERNAL STANDARD(S)
MODELS
LOCATES MODEL PEAKS
CLEANUP
EXTRACTS SPECTRA FROM RAW DATA
PBM UNK
CONDENSATION OF UNKNOWN SPECTRA FOR INPUT TO PBM
PBM
REVERSE SEARCH MATCHING PROGRAM
RRTCON
RRT AND CONCENTRATION CALCULATION
HISLIB
TENTATIVE IDENTIFICATION OF SPECTRA AND UPDATE OF HISTORICAL DATABASE

## Create Control File (4A)

At user terminal,

```
            PDS> LOGIN LYN

            PASSWORD?

            PDS> INE DO CONTROL

              OPTIONS (1) TAPE NO

                      (2) RUN NAMES

                      (3) EXIT

         (1) ENTER OPTION ?  1

             STANDARD TAPE (RUN, DELETE) Y OR N ?

              <If Yes, the defaults are used.  All
              modules   are   run,   and   all   internal
              files are deleted (See above).
              If No, the user is prompted as to whether
              each module is to be run.>

             ENTER TAPE NO ?  <enter tape number (TAPE.NO)>

             ENTER MIN/MAX TIC VALUE ?  <enter tic
             value.  At this time TAPE.LOG data-
             base is updated with this value>

             Control now returns to OPTIONS.

         (2) ENTER OPTION ?  2

             ENTER RUN NAME ?  <enter desired run name>

             STANDARD RUN (RUN,DELETE) Y OR N ?

              <If yes, defaults are used.  (See above)
              If no, user is prompted for each flag.>

             ENTER MIN/MAX TIC VALUE ?  <enter tic value>

             ENTER RUN NAME ?  < enter run name or hit CR to end>

             When <CR> is entered control returns to OPTIONS.
```

```
(3) ENTER OPTION ?  1

    PDS> LOGO
```

This procedure creates a CONTROL.DAT file, such as the following

```
  6937 347 19   1000   1000ACISTD000000000110
D10 ANTHRACENE      20                                    2000288
NUS                      TENAX GC
  6949 347 12   1000   1000B/NSTD000000000110
D10 ANTHRACENE      20                                    2000288
NUS                      1% SP-2250
  6950 347  1   1000   1000B/NSTD000000000110
D10 ANTHRACENE      20                                    2000288
NUS                      1% SP-2250
```

## CREATE AND SUBMIT BATCH FILE (4B AND 4C)

See Appendix C for file structure.

The program BATCHJOB is run by the GC/MS user to build a series of batch jobs which execute various tasks associated with the GC/MS system. The job series created consists of a CONTROL.BIS file, first job of the series, BATCHnnn.BIS files which execute various GC/MS program(s), and a DELETE.BIS file, last of the the series, which deletes all the .BIS files created. All batch files of the series are linked in that CONTROL.BIS submits BATCH001.BIS which submits BATCHnnn.BIS which submits DELETE.BIS. nnn is initially set to '001' and is incremented by 1 for each run to be processed.

The BATCHJOB program uses the file CONTROL.DAT created by the INFORM procedure outlined above to create the series of batch jobs. After running the CONTROL INFORM procedure, the user should run the BATCHJOB program which executes the following dialogue.

On EGD account:

```
PDS> RUN BATCHJOB

ENTER INPUT FILE NAME>

        <Reply with name of file created by the INFORM
         CONTROL procedure, normally CONTROL.DAT.>

ENTER USER ACCOUNT NAME>
```

<Reply with user account name used for GC/MS operations, normally EGD.>

ENTER LINEAR RATIO> (SEE REFERENCE 2)

<Enter value to be used by analysis programs.>

ENTER MAGTAPE UNIT NUMBER (0 OR 1) >

<Enter number of drive where the converted tape with the appropriate data is mounted.>

At this point BATCHJOB will create the series of batch job files. The following describes the batch files created by this program to execute the runs specified by CONTROL.DAT.

The CONTROL.BIS file contains commands to perform the following tasks:

1. Deletes various work files.
2. Creates the file UNITS.DAT to assign I/O units.
3. The following is done for each set of data in the CONTROL.DAT file:
    a. Requests tape mount of appropriate tape (if necessary).
    b. Creates the file FOR005.DAT using constant information (see BATCHnnn.BIS descriptions) plus information created by the CONTROL procedure form the CONTROL.DAT file.
    c. Creates files named in the CONTROL.DAT file.
    d. Executes the RETRIEVE program.
4. Rewinds the tape.
5. Submits the file BATCH001.BIS.

For each set of data specified in the CONTROL.DAT file, a BATCHnnn.BIS file is created. Each BATCHnnn.BIS file performs the following tasks:

1. Creates a FOR006.DAT file containing answers to the CLEANUP program queries (see Appendix C for file format). Data from the CONTROL.DAT file is used to create records for this file.
2. According to the flags contained in CONTROL.DAT (see page 25 for detailed description) the following programs runs are included in the batch job: BUILD, STANDARD, PEAK, MODELS, CLEAN, SPECTRA, PBMUNK, TIMSEK, SPECIAL, and HISLIB.
3. Submit the next batch job file, or, on the last one, submit the DELETE.BIS file.

The DELETE.BIS file performs the following tasks:

1. Deletes CONTROL.BIS file.
2. Deletes each of the BATCHnnn.BIS files.
3. Deletes the DELETE.BIS file.

Once BATCHJOB has successfully completed, mount the appropriate converted tape on the tape drive specified in BATCHJOB and submit CONTROL.

```
PDS> SUB CONTROL
```

# Computer Analysis Modules (5)

(See Appendix D for graphic descriptions of inputs and outputs for each module.)

## A. RETRIEVE

This program was designed to be executed in the batch stream and its main function is to transfer a user-specified file from tape to the requested device and file.

NOTE: The output device and file name are generally specificed by: DB0:<run name>.FIL

NOTE: These files may be retrieved from tape interactively by using the following procedure. The file position on the tape must be known before the file may be retrieved. The first command should be to position the tape at the correct file. This is accomplished by skipping <file position -1> number of files. After the tape has been positioned, the file can be read with the READ command.

## B. BUILD

The module BUILD creates <Runname>.DAT, <Runname>.PAR, the files that are required by the modules STANDARD, PEAK, MODELS, and CLEAN. The output files created can be located on the device as specified during input to the module. The following command will initiate the program:

```
PDS> RUN BUILD
```

## C. STANDARD

This module performs a Rindfleisch Background Correction Algorithm (4) on 7-8 selected masses for each internal standard. This program requires the files created by BUILD for input. The following command will initiate execution:

```
PDS> RUN STANDARD
```

### D. PEAK

PEAK is used to provide a means of locating a scan where 5-8 specific masses, corrected for background by STANDARD, all maximize simultaneously. STANDARD and PEAK are used together for locating the internal standard(s). Therefore, STANDARD must be executed prior to PEAK. The following command will initiate execution.

```
PDS> RUN PEAK
```

### E. MODELS

To execute this program interactively, the module BUILD must have been run. This module locates model peaks (peak height and shape) that are used by the module CLEANUP. The following command will start the program MODELS:

```
PDS> RUN MODELS
```

### F. CLEAN-UP

This module, in combination with MODELS, is used to extract spectra from each GC/MS run. The interactive use of this program requires that BUILD and MODELS have been executed. The following command will initiate execution.

```
PDS> RUN CLEAN
```

### G. PBM (3)

The Probability Based Matching (PBM) system is used to match unknown spectra extracted by CLEAN against a spectrum library. To execute this program the following commands must be issued:

```
PDS> COPY DB0:<Run Name>.CLN DB1:EOR002.DAT

PDS> RUN PBMUNK

PDS> RUN PBM
```

To add spectra to the PBM spectrum library, see Appendix E.

## H. TIMSEK

This program calculates the retention time and concentration for each component relative to the internal standard(s). Output from the following modules must be present: RETRIEVE, PEAK, CLEAN-UP, and PBM. The following command will initiate execution.

```
PDS> RUN TIMSEK
```

## I. HISLIB

The Historical Library Program (HISLIB) is designed to provide tentative identification of spectra through the use of RRT's calculated by the TIMSEK program and the compound matches provided by PBM. The program (HISLIB) is a two step process for the identification of spectra extracted from the GC/MS run:

Pass 1 - Automated tentative identification of spectra using the output created by TIMSEK.
Pass 2 - Interactive processing of the chemist verified spectra not identified in Pass 1.

Spectra tentatively identified by Pass 1 of HISLIB are stored in a transaction file, TRANSACT.HIT. Spectra considered as misses (those with no acceptable PBM matches) are stored in the transaction file, TRANSACT.MIS. These files allow for multiple batch streams in the analysis of GC/MS runs and provide time to identify and correct any problems that may have occured during Pass 1 of HISLIB. To update this information into the HISLIB.HIT and HISLIB.MIS databases, the following batch job is submitted.

```
TRANSACT.BIS
$JOB EGD TRANSACT 9999
$APP [203,377]TRANSACT.HIT TRANSACT.HIT
$APP [203,204]TRANSACT.HIT TRANSACT.HIT
$COPY TRANSACT.HIT HIT.TMP
$APP [203,377]TRANSACT.MIS TRANSACT.MIS
$APP [203,204]TRANSACT.MIS TRANSACT.MIS
$COPY TRANSACT.MIS  MISS.TMP
$SUB TRANSACT1
$SUB TRANSACT2
$COPY HISLIB.HIT [203,377]*.*
$COPY RRT.HIT [203,377]*.*
$RUN [203,377]TRANSHIT
$COPY GCWINDOW.MIS [203,377]*.*
$COPY RRT.MIS [203,377]*.*
$COPY HISLIB.MIS [203,377]*.*
$RUN [203,377]TRANSMIS
$EOJ
```

This job collects transaction files from all accounts where HISLIB may have been processed, creates backup copies of these files and the HISLIB databases, creates new transaction files for subsequent HISLIB processing, and runs TRANSHIT and TRANSMIS to update the HIT and MIS databases, respectively.

Limitations associated with the Historical Library Program include:

1) Only one run at a time may be processed through the first pass of the Historical Library Program on an individual account. (There may be more than one account using the Historical Library Program simultaneously. A separate transaction file is established on each account and these files can be accessed with write privileges by only one program at a time.)

2) The transaction file can not be processed while a run is being processed through the Historical Library.

3) A run should not be processed through the second pass of the Historical Library until the transaction file, which contains matches identified during the first pass, has been processed. This procedure avoids the possibility of adding a compound to the Historical Database that has been identified during the first pass.

## HISTORICAL LIBRARY ALGORITHM

(Automated Processing - Pass 1)

Each spectrum extracted from the GC/MS run by CLEANUP is compared with each compound RRT entry in the Historical Library. Each RRT entry is uniquely identified by its CAS#, GC Column, and internal standard. A comparison of the unknown spectrum's RRT'(s) with the Library's RRT entries is accomplished in the following steps (see accompanying diagram).

1. The "best" PBM match is found for the present unknown spectrum. For a PBM match to be considered it must meet the following criteria.

   K value / (K value + Delta K) > .44

   and Percent contamination < 70

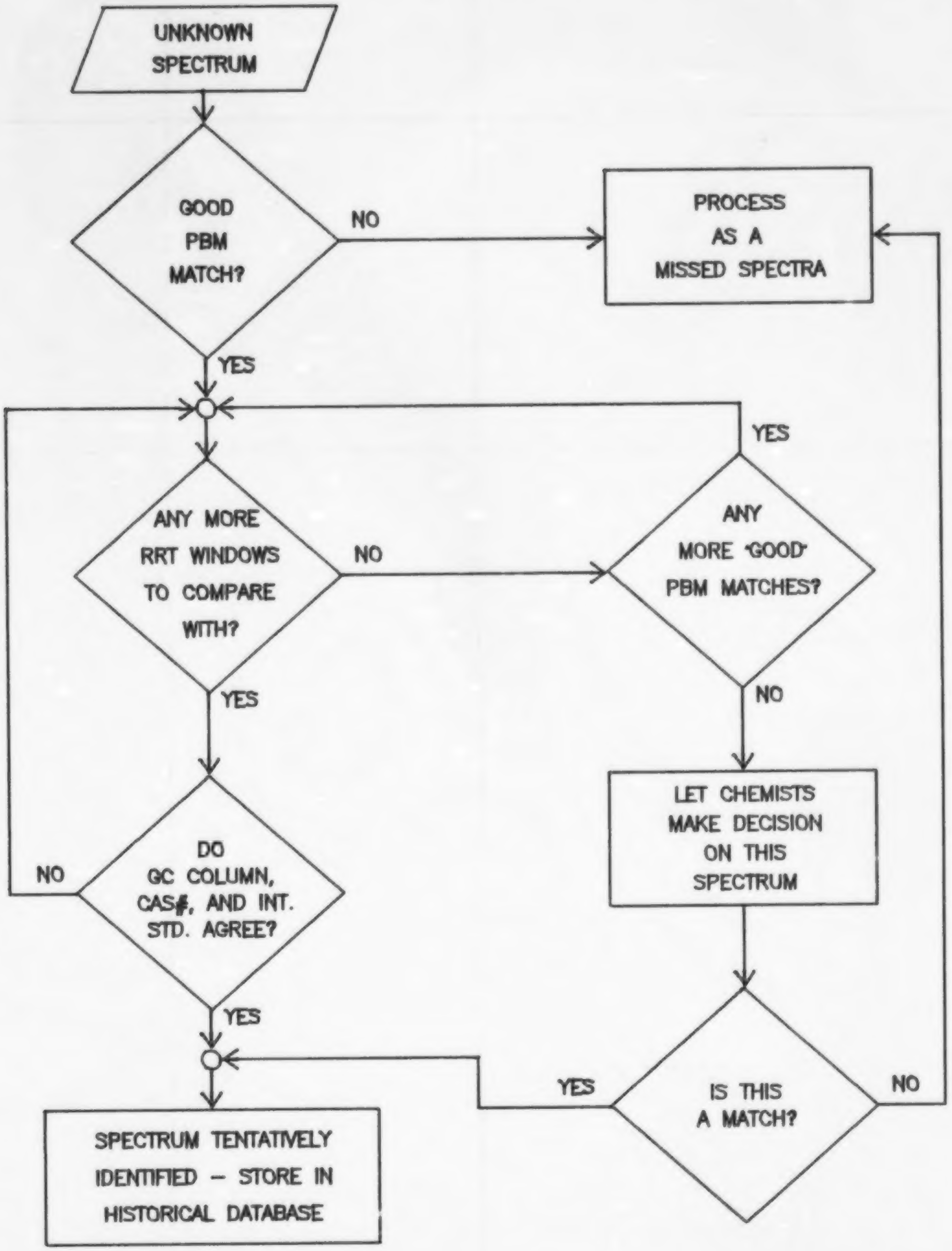
   where K, Delta K, and percent contamination are PBM match parameters. The "best" match is that with the highest value for K value / (K value + Delta K). If no matches are found that meet the above criteria, the spectrum is processed as a "missed" spectrum.

***NOTE - Because all spectra in the PBM library do not contain CAS #'s, a "best" match without a CAS # is placed in the .HLD file for chemist review.

2. The RRT, Cas #, GC column and internal standard for the "best" match are compared with each entry in the historical library. If the RRT falls within the allowable window and the other values match, this compound is confirmed as a tentative identification, or "hit". The historical library is updated and the next unknown spectrum is processed.

3. If the "best" PBM match is not identified by comparison of RRT, Cas #, internal standard, and GC column, then the "next best" PBM match is located (as in Step 1) and compared against the Historical Library (as in Step 2). If there are no more "next best" PBM matches, then this spectrum is placed in the .HLD file for chemist review, and the next unknown spectrum is automatically processed.

# HISLIB LOGIC


UNKNOWN SPECTRUM
GOOD PBM MATCH?
NO
PROCESS AS A MISSED SPECTRA
YES
ANY MORE RRT WINDOWS TO COMPARE WITH?
NO
YES
ANY MORE 'GOOD' PBM MATCHES?
YES
NO
DO GC COLUMN, CAS#, AND INT. STD. AGREE?
NO
YES
LET CHEMISTS MAKE DECISION ON THIS SPECTRUM
IS THIS A MATCH?
YES
NO
SPECTRUM TENTATIVELY IDENTIFIED – STORE IN HISTORICAL DATABASE

(Interactive Use of Hislib - Pass 2)

The interactive use of HISLIB to process .HLD files is initiated by the following commands:

```
PDS> LOGIN CHEM
PASSWORD?
```

The file UNITS.DAT must be set correctly for interactive use of the Historical Library program (see page 27).

```
PDS> RUN HISLIB

ENTER RUN NAME> (Enter appropriate run name)

ENTER DEVICE> DB0:

ENTER STARTING SCAN> (Enter first scan to be
processed. CR defaults to first scan of .HLD file.)

ENTER ENDING SCAN> (Enter last scan to be processed.
CR defaults to last scan of .HLD file.)
```

To view one scan only, enter that scan number for both starting and ending scan. STOP exits from HISLIB.

The user is then prompted to process each requested scan of the given run as follows:

```
REJECT THIS SPECTRUM? (Enter Y or N)
```

If answered Y, the program will proceed to the next scan. A response of N gives the prompt:

```
DOES SPECTRUM BELONG IN MISS FILE? (Enter Y or N)
```

If answered Y, the next scan is prompted, and if answered N the following choice is given.

```
DOES SPECTRUM BELONG IN HIT FILE? (Answer Y or N)
```

If N is answered to all three prompts, the scan is repeated. Upon answering Y to the last choice, the 10 best PBM matches, their names, compound #'s, and CAS#'s (if available) are printed on the screen and the following is prompted.

```
ENTER COMPOUND NUMBER> (Enter appropriate number.)
```

Compound # 0 always corresponds to no entry, and this choice sends the scan to the MISS file and the program proceeds to the next scan. Figure IV shows an example of a new compound entered into HISLIB. After entering the compound number, the user is told:

```
YOU ENTERED BENZOPHENONE
IS THIS CORRECT?  (Enter Y or N)
```

If answered N, the choices are repeated. If answered Y, the following prompt is given:

```
IS THIS A NEW COMPOUND (OR RRT WINDOW)?  (Enter Y or N)
```

If Y is entered, the user is given the options of changing the RRT window and/or entering more compounds from the scan:

```
DO YOU WISH TO CHANGE RRT?  (Enter Y or N)
MORE COMPOUNDS TO ENTER FROM THIS SCAN?  (Enter Y or N)
```

If a compound already in RRT.HIT (for that particular column and internal standard) is entered and specified as a new compound, an error message is printed (Fig. V).

```
***COMPOUND RRT ALREADY PRESENT - OUTSIDE RRT WINDOW***
DO YOU WANT TO ADD THIS COMPOUND TO DATABASE?  (Enter Y
or N)
```

If answered N, the compound choices are repeated. An answer or Y adds the compound to the database without creating a new RRT window.

If a compound that is not new (already in RRT.HIT for that particular column and internal standard) is chosen, a response of N to

```
IS THIS A NEW COMPOUND (OR RRT WINDOW)?  (Enter Y or N)
```

updates the database without creating a new window. If N is answered to this prompt, but the compound is in fact new, the following error message is printed (Fig. VI).

```
***ERROR - COMPOUND NOT MATCHED   (PLEASE REENTER)***
```

and the list of choices is repeated.

In addition to checking for new and existing entries, the HISLIB program also prevents duplicate entries of compounds for a given run (Fig. VII). If a compound is chosen for which there is no CAS#, the user is prompted to enter the CAS# after declaring that the compound choice is correct. This number must be entered before the program will proceed, and the corresponding PBM Record # is updated.

Occasionally, the chemist is able to identify a spectrum in the .HLD file as a compound that already exists in HISLIB, but has not been listed as one of the 10 best PBM matches. In this case, the last option (chemist manual entry) is chosen. The user is prompted for the CAS number and asked to verify its accuracy:

```
     ENTER CAS NUMBER> 108883
     IS 108883 CORRECT?  (Enter Y or N)
```

If answered N, the user is again prompted for the CAS number. If answered Y, the user is informed of the compound selected:

```
     THE COMPOUND YOU HAVE SELECTED IS:  TOLUENE

     IS THIS CORRECT?  (Enter Y or N)
```

If answered Y, the compound is updated in HISLIB. If answered N, the choices are repeated. If the CAS number entered does not correspond to a compound that exists in HISLIB for the GC column of interest, the following message is printed:

```
     ***ERROR-COMPOUND NOT MATCHED (PLEASE REENTER)***
```

and the choices are repeated.

## FIGURE IV

```
RUN - 946     SCAN - 385
REJECT THIS SPECTRUM? N
DOES SPECTRUM BELONG IN MISS FILE? N
DOES SPECTRUM BELONG IN HIT FILE? Y

   Cmpd     PBM       CAS
     #     Rec #       #        Compound Name

     0                        No Entry
     1     24255         0    ETHYL BENZOATE
     2      6988    119619    BENZOPHENONE
     3     30437     98862    METHYL-PHENYL KETONE
     4     24265     92875    BENZIDINE
     5     25446    119619    BENZOPHENONE
     6      3080         0    BENZOIC ACID HYDRAZIDE
     7     31439     93583    METHYL BENZOATE
     8     20014    119619    BENZOPHENONE
     9      8986    134816    BENZIL
    10     24201    119619    BENZOPHENONE
    11                        CHEMIST MANUAL ENTRY

ENTER COMPOUND NUMBER> 10
YOU ENTERED BENZOPHENONE
IS THIS CORRECT?> Y

IS THIS A NEW COMPOUND (OR RRT WINDOW)? Y
DO YOU WISH TO CHANGE RRT? (Y or N)> N
MORE COMPOUNDS TO ENTER FROM THIS SCAN? N
```

FIGURE V

```
RUN - 5314     SCAN - 117
REJECT THIS SPECTRUM? N
DOES SPECTRUM BELONG IN MISS FILE? N
DOES SPECTRUM BELONG IN HIT FILE? Y

   Cmpd    PBM      CAS
     #     Rec #      #        Compound Name

     0                       No Entry
     1      1897    67663   TRICHLOROMETHANE (CHLOROFORM)
     2      5476    75274   BROMODICHLOROMETHANE
     3     17432   144627   OXALIC ACID
     4     18286    67663   CHLOROFORM
     5     18227    67663   TRICHLOROMETHANE (CHLOROFORM)
     6     30244   254212   1,1-DIFLUORO-1,2,2-TRICHLOROETHANE
     7                       CHEMIST MANUAL ENTRY

ENTER COMPOUND NUMBER> 4
YOU ENTERED CHLOROFORM
IS THIS CORRECT?> Y

IS THIS A NEW COMPOUND (OR RRT WINDOW)? Y

***Compound RRT already present - outside RRT window***

DO YOU WANT TO ADD THIS COMPOUND TO DATABASE? Y
```

**FIGURE VI**

```
RUN - 925      SCAN - 140
REJECT THIS SPECTRUM? N
DOES SPECTRUM BELONG IN MISS FILE? N
DOES SPECTRUM BELONG IN HIT FILE? Y

   Cmpd     PBM       CAS
     #      Rec #      #         Compound Name

     0                        No Entry
     1       2994    99832    ALPHA PHELLANDRENE
     2      28374   527844    1-METHYL-2-ISOPROPYLBENZENE
     3       2838   535773    1-METHYL-3-ISOPROPYLBENZENE
     4       2855    99876    1-METHYL-4-ISOPROPYLBENZENE
     5      18725   527844    O-CYMENE
     6       2855  1074437    M-N-PROPYLTOLUENE
     7      18727    99876    P-CYMENE
     8      18726   535773    M-CYMENE
     9      18729    98776    P-CYMENE
    10       2847  1758889    1,4-DIMETHYL-2-ETHYLBENZENE
    11                        CHEMIST MANUAL ENTRY

ENTER COMPOUND NUMBER> 5
YOU ENTERED O-CYMENE
IS THIS CORRECT?> Y

IS THIS A NEW COMPOUND (OR RRT WINDOW)? N

***ERROR - COMPOUND NOT MATCHED (PLEASE REENTER)***
```

## FIGURE VII

```
RUN - 64     SCAN - 342
REJECT THIS SPECTRUM? N
DOES SPECTRUM BELONG IN MISS FILE? N
DOES SPECTRUM BELONG IN HIT FILE? Y

   Cmpd     PBM       CAS
     #     Rec #       #        Compound Name

     0                         No Entry
     1     32579   108883     TOLUENE
     2       655   108883     METHYLBENZENE (TOLUENE)
     3       656   544252     1,3,5-CYCLOHEPTATRIENE
     4     25767        0     ALLYLIDENE CYCLOBUTENE
     5     25766        0     DL-SPIRO(3.3)HEPTA-1,5-DIENE
     6     17435   544252     CYCLOHEPTATRIENE
     7     17436   108883     METHYLBENZENE (TOLUENE)
     8     17437   108883     METHYLBENZENE (TOLUENE)
     9     17438   108883     TOLUENE
    10     17439   108883     1,4-DIMETHYL-2-ETHYLBENZENE
    11                         CHEMIST MANUAL ENTRY

ENTER COMPOUND NUMBER> 10
YOU ENTERED METHYLBENZENE (TOLUENE)
IS THIS CORRECT?> Y

IS THIS A NEW COMPOUND (OR RRT WINDOW)? N

***ERROR - COMPOUND ALREADY ENTERED FOR THIS RUN***

RUN - 64     SCAN - 338
REJECT THIS SPECTRUM?
```

FIGURE VIII

```
RUN - 4823     SCAN - 286
REJECT THIS SPECTRUM? N
DOES SPECTRUM BELONG IN MISS FILE? N
DOES SPECTRUM BELONG IN HIT FILE? Y

   Cmpd     PBM       CAS
     #     Rec #       #       Compound Name

    0                         No Entry
    1      3086  13269528    TRANS-3-HEXENE
    2      4213    760203    1-PENTENE,3-METHYL
    3      5526  14686136    TRANS-2-HEPTENE
    4      4034    592416    1-HEXENE
    5      5681  14686136    2-HEPTENE,(E)
    6      3515    592767    1-HEPTENE
    7      2974    111273    1-HEXANOL
    8      4141   7385786    1-PENTENE,3,4-DIMETHYL
    9      5772  53535334    HEPTANOL(VAN9CI)
   10      4155    558372    3,3-DIMETHYL-1-BUTENE
   11                        CHEMIST MANUAL ENTRY

ENTER COMPOUND NUMBER> 11
ENTER CAS NUMBER> 108883
IS 108883 CORRECT?> Y
THE COMPOUND YOU HAVE SELECTED IS:  TOLUENE
IS THIS CORRECT?  Y
```

## Graphics (6)

The GC/MS graphics package is described in this section. Both interactive and batch plotting capabilities are available. The interactive program, LYN, accesses raw, library, and cleaned-up spectra. SPECTRA3 batch plots library spectra and SPECTRA4 batch plots cleaned-up spectra. The formats available for the interactive program and their associated file specifications are as follows:

| FORMAT | FILE TYPE | DESCRIPTION |
|---|---|---|
| RAW | <EXPLICIT.DAT> | raw data generated by BUILD |
|  | <EXPLICIT.BCK> | raw data generated by STANDARD |
| CLN | <EXPLICIT.CLN> | cleaned-up spectra generated by CLEAN |
|  | <EXPLICIT.SEK> | cleaned-up spectra generated by TIMSEK |
|  | <EXPLICIT.HLD> | cleaned-up spectra generated by Pass 1 of HISLIB |
| LIB | NAMES2.DIR | library spectra |

To run the interactive program, the following commands are issued from a Tektronix 4014 terminal:

```
PDS> LOGIN CHEM

PASSWORD?

PDS> RUN LYN

OPENING FILE> DB0:[203,203]NAMES2.DIR
```

The file of library spectra, NAMES2.DIR, is automatically opened. To access raw or cleaned-up data files, a format statement is required. For example, the raw file DB0:[203,203]3989.DAT would be opened as follows.

```
Lyn > FOR RAW

Lyn > FILE DB0:[203,203]3989.DAT

OPENING FILE > DB0:[203,203]3989.DAT
```

```
Summary for file > DB0:[203,203]3989.DAT
Low mass        45
High mass      455
Scans          399
Mass range     411
```

Files that have been opened can be accessed interchangeably by a TYPE statement. For example,

```
Lyn > TYPE LIB
```

returns to library spectra after opening raw or cleaned-up files. The last opened file or the last file type specified becomes the default file.

The graphics package provides mass spectra and reconstructed gas chromatograms (RGC's). Mass spectra can be plotted with the SPEC command or listed with the LIST command. These are summarized below.

## SPEC AND LIST COMMANDS

| COMMAND | RESULT |
|---|---|
| SPEC <CR> | Outputs bar plot of default spectrum # over default mass range. |
| SPEC XXXX <CR> | Outputs bar plot of spectrum # XXXX over default mass range and establishes XXXX as new default value. |
| SPEC XXXX/XXXX,XXXX | Outputs bar plot of spectrum # XXXX over mass range XXXX,XXXX. Establishes new default spectrum # and mass range. |
| SPEC XXXX-XXXX | Perform spectrum subtraction XXXX-XXXX and outputs the results bar plot over the default mass range. |
| SPEC XXXX-XXXX/XXXX,XXXX | Perform spectrum subtraction XXXX-XXXX and outputs resultant bar plot over mass range XXXX-XXXX. |
| LIST <CR> | Outputs listing of mass intensity pairs for default spectrum over default mass range. |
| LIST XXXX | Outputs listing of spectrum XXXX over default mass range. |

| | |
|---|---|
| LIST XXXX/XXXX,XXXX | Outputs listing of spectrum XXXX over mass range XXXX,XXXX. |
| LIST XXXX-XXXX/XXXX,XXXX | Outputs listing of mass intensity pairs for default spectrum over default mass range. |

The command RGC is used to plot reconstructed gas chromatograms; it operates only on RAW data files. The following summarizes the capabilities of the RGC command:

## RGC COMMANDS

| COMMAND | RESULT |
|---|---|
| RGC <CR> | Graphs total ion current for default mass range over default scan range. |
| RGC XXXX | Graphs extracted ion current profile (EICP) for mass XXXX over default scan range. |
| RGC XXXX/XXXX,XXXX | Graphs EICP for specified mass over specified scan range. |
| RGC XXXX,XXXX | Graphs ion current profile summarized for mass range XXXX,XXXX over default scan range. |
| RGC XXXX,XXXX/XXXX,XXXX | Graphs ion current profile summarized for mass range XXXX,XXXX over scan range XXXX,XXXX. |

Up to three RGC's can be requested in one command if separated by semicolons. For example,

```
RGC 188;189;187/300,400
```

graphs EICP's for masses 188, 189, and 187 over the scan range 300,400.

Areas can be calculated for raw data files using the AREA command. Mass and scan ranges are specified as in RGC. For example,

```
AREA 188/275,285
```

calculates the area for mass 188 from scan 275 to scan 285. In addition, the command STATUS will list current default parameters, and STOP, EXIT, or QUIT exit from the program.

To use the program SPECTRA3 for batch plotting of library spectra, create a batch input file as follows:

```
$JOB PBM SPECTRA3 9999
$CREATE UNITERS.DAT
4 0
$EOD
$ASS DB1:  4
$CREATE FOR004.DAT
[203,203]NAMES2
DB0   2282
  2283
 11757
 24473
 24477
 24475
 24367
  5406
 31632
 10198
 24476
     0
$RUN SPECTRA3
$DEL SPECTRA3.TSK;*
$PLOT DB0:NAMES2 1.0
$DEL UNITERS.DAT;*
$DEL FOR004.DAT;*
$EOJ
```

After "DBO", the serial numbers of library spectra to be plotted are listed in I6 format, right justified. The last serial number is followed by a zero in column 6.

To plot cleaned-up spectra from .HLD files with SPECTRA4, a batch input file is created as follows.

```
$JOB PBM SPECTRA4 9999
$CREATE UNITERS.DAT
4 0
$ASS DB1:  4
$CREATE FOR004.DAT
[203,204]4141
DBO
231
274
280
282
300
342
357
364
  0
$EOD
$RUN SPECTRA4
$DEL SPECTRA4.TSK;*
$PLOT DBO:[203,204]4141 1.0
$DEL UNITERS.DAT;*
$DEL FOR004.DAT;*
$EOJ
```

In this case, file [203,204]4141.HLD is accessed. The scans to be plotted are listed in I3 format, right justified, following the line "DBO". After all scans are listed a zero is placed in column 3. To plot spectra from other .HLD files, the section from "$CREATE FOR004.DAT" to "$PLOT DBO:[203,204]4141 1.0" is repeated with the appropriate file name and scan numbers.

Report Generation (7)

Four report programs are available to summarize data from the .HIT databases:

1) MONTHLY lists all entries in RRT.HIT, i.e., the unique compounds entered for each GC column

2) WEEKLY reports all entries made to HISLIB.HIT during a given time interval

3) FREQHISLB summarizes the total number of occurrences for each unique compound in HISLIB in order from highest to lowest frequency

4) FREQIC provides breakdowns of FREQHISLB by industrial category, first as a list of all unique compounds tentatively identified for a given industry, and second as a list of all industries for which a given compound has been tentatively identified.

One program, MISREPORT, summarizes all data in the .MIS databases.

To generate the MONTHLY or WEEKLY report, the program HITS on the EGD account is run. The user then enters MONTHLY or WEEKLY as desired. Each is described in detail below:

**MONTHLY:** When generating monthly reports, all the data in RRT.HIT is included, and the user does not specify a time interval. Only the GC columns to be included and fields for sorting are user selected.

The user is first asked

>Are all GC columns to be included? *(Enter Y or N)*

If "YES", this is the end of this step. If "NO", the user is prompted to enter the proper code from Figure IX. The user is then asked

>Do you want to include another GC column? *(Enter Y or N)*

If "NO", this is the end of this step. If "YES", the user is again prompted for a code. Only two GC columns are allowed, and after entering the second code the user is asked to enter sorting codes for up to 5 fields in I2 format. (See Figure X for sorting codes for monthly report. A "0" or "control Z" is entered after all codes have been entered. Output from this program is placed in

DB0:MONTHLY.DAT.

**WEEKLY:** This report extracts all the information in HISLIB.HIT for a given fraction(s) and a specified time interval. The user is first asked to

```
>Enter types of runs to be included:
```

The appropriate code is entered from the following list of choices:

1 - all runs
2 - only blanks
3 - only spikes
4 - only standards
5 - only QC's
6 - only non-blank, non-spike, non-standard, non-QC and non-CRS
7 - only CRS

The user is then prompted to enter those GC columns to be included

```
>Are all GC columns to be included? (Enter Y or N)
```

This section is completed in the same way as for the MONTHLY. The next prompts are

```
>ENTER BEGINNING DATE:
>ENTER ENDING DATE:
```

These are of the format "21-DEC-80". Lastly, the user is asked to enter codes as for the MONTHLY, using the sorting codes for weekly reports in Figure X. Output for this report is placed in DB0:WEEKLY.DAT.

**FREQHISLB & FREQIC:** For the FREQHISLB and FREQIC reports, no user inputs are required. They automatically include all non-blank, non-spike, non-standard, non-QC, and non-CRS data entered into HISLIB to date. These programs are initiated by the commands

```
INF DO FREQHISLB
```
and
```
INF DO FREQIC
```

for which the user must be logged into the account LYN.

Output for FREQHISLB will be found in DB0:FREQHISLB.DAT. The program FREQIC creates two output files: 1) DB0:FREQIC1.DAT is a list of all compounds in HISLIB broken down by their frequency in the various industries; and 2) DB0:FREQIC2.DAT is a list of the compounds tentatively identified in each industrial category in order of frequency.

**MISREPORT:** Two options are available for the MISREPORT program (on PBM account). Report 1 summarizes all entries for one or more GC columns. Report 2 lists all matches for a given entry (unknown component).

The user is first asked to enter the desired report number (I1 format). If '1' is selected, the next prompt is

```
>Enter minimum frequency value (I3):
```

The report will include all entries occurring with frequency greater than or equal to this value. If '001' is entered, all entries will be reported. The user is then asked to

```
>Enter concentration value (F10.0)
```

Only entries containing at least on occurrence with concentration greater than or equal to this value will be reported. The user is then asked to

```
>Enter # of GC columns to use (I3)
```

If '000' is entered, all GC columns are used. If any other value is entered, the user is prompted to enter the appropriate code(s) from Table IX (I3). Output for this program will be found in DB0:FOR006.DAT.

For Report 2, the user is asked to enter the GC column code from Table IX (I2), the run name (6A1) and the scan number (I3) for the entry to be reported. Output for this program is also found in DB0:FOR006.DAT.

**FIGURE IX**

| GC COLUMN | CODE |
|---|---|
| 3% OV1 | 01 |
| Tenax GC | 02 |
| 1% SP-2250 | 03 |
| 0.2% Carbowax 1500 Carbopak A | 04 |
| 3% SP2250 DB PESTICIDE | 05 |
| 0.2% Carbowax 1500 Carbopak C | 06 |
| 1% SP-1240 DA | 07 |
| Ultrabond 20 M | 08 |
| 3% SP-2250 DB | 09 |
| 0.3% Carbowax 20M | 10 |

## FIGURE X

### SORTING CODES FOR MONTHLY REPORTS

| CODE | FIELD |
|---|---|
| 1 | Relative Retention Time |
| 2 | Cas # |
| 3 | Compound Name |
| 4 | Relative Conc. |
| 5 | Anal. Lab Code |
| 6 | Industrial Code |
| 7 | GC Col. Code |
| 8 | Run Name |
| 9 | Scan # |
| 10 | Internal Standards |
| 11 | Fraction Type |
| 12 | Record # in RRT.HIT file |
| 13 | Entry # in RRT.HIT file |

### SORTING CODES FOR WEEKLY REPORTS

| CODE | FIELD |
|---|---|
| 1 | Window Relative Retention Time |
| 2 | Actual R. R. Time |
| 3 | Compound Name |
| 4 | Run Name |
| 5 | Scan # |
| 6 | Industrial Code |
| 7 | GC Col. Code |
| 8 | Anal. Lab Code |
| 9 | Internal Standard Code |
| 10 | Relative Conc. |
| 11 | Fraction Type |
| 12 | Cas # |

# REFERENCES

(1) DENNIS H. SMITH, MICHAEL ACHENBACH, WILLIAM J. YEAGER, PATRICIA J. ANDERSON, WILLIAM L. FITCH, AND THOMAS C. RINDFLEISCH, "QUANTITATIVE COMPARISON OF COMBINED GAS CHROMATOGRAPHIC/MASS SPECTROMETRIC PROFILES OF COMPLEX MIXTURES," ANALYTICAL CHEMISTRY, VOL. 49, NO. 11, PAGE 1623, SEPT. 1977.

(2) R. G. DROMEY, MARK J. STEFIK, THOMAS C. RINDFLEISCH, AND ALAN M. DUFFIELD, "EXTRACTION OF MASS SPECTRA FREE OF BACKGROUND AND NEIGHBORING COMPONENT CONTRIBUTIONS FROM GAS CHROMATOGRAPHY/MASS SPECTROMETRY DATA," ANALYTICAL CHEMISTRY, VOL. 48, NO. 9, PAGE 1371, AUGUST 1976.

(3) GAIL M. PESYNA, RENGACHARI VENKATARAGHAVAN, HENRY E. DAYRINGER, AND F. W. MCLAFFERTY, "PROBABILITY BASED MATCHING SYSTEM USING A LARGE COLLECTION OF REFERENCE MASS SPECTRA," ANALYTICAL CHEMISTRY, VOL. 48, NO. 9, PAGE 1362, AUGUST 1976.

(4) R.E. SUMMONS, W.E. PERIERA, W.E. REYNOLDS, T.C. RINDFLEISCH AND A.M. DUFFIELD, "ANALYSIS OF TWELVE AMINO ACIDS IN BIOLOGICAL FLUIDS BY MASS FRAGMENTOGRAPHY," ANALYTICAL CHEMISTRY, VOL. 46, NO. 4, PAGE 582-587, 1974.

## INFORM Data Base (TAPE.LOG) Definition

| FIELD | DESCRIPTION |
|---|---|
| RUN.NAME | AERL assigned identification unique for each GC/MS run |
| TAPE.NO | AERL assigned TAPE.NO |
| EXTRACT.NO | Indicates site possessing extract (AERL, RTI, etc.) |
| EXTRACT.LOC | AERL assigned code describing location |
| USER.RUN.NAME | User file name |
| USER FILE.NO. | Order in which run falls on tape |
| USER.SAMPLE.NO | Lab sample no. |
| FRACTION.TYPE | First 3 characters indicate physical separation that was made, i.e., VOA ACI, B/N, PES, DAI. Last 3 characters indicate blanks (BLK), standards (STD), spikes, (SPK), etc. |
| INDUSTRIAL.CAT | Industry type of sample source 49 different categories |
| STAGE.TREATMENT | Point which sample is taken |
| DATE.SAMPLED | Sample date |
| DATE.RUN | Date GC/MS analysis is run |

| | |
|---|---|
| SAMPLER | Contractor who does sampling |
| MS | Brand of MS |
| GC | Brand of GC |
| DATA.SYSTEM | Data system used, i.e., INCOS, SYSTEM 150, HP, VARIAN |
| AERL.TAPE.NO | Tape# of internal conversion tape |
| AERL.FILE.NO | File location on converted tapes |
| CONC.FACTOR | Concentration factor of extract |
| INJ.VOLUME | Volume of injection |
| INT.STD.1.CONC | Concentration of internal standard #1 |
| INT.STD.2.CONC. | Concentration of internal standard #2 |
| INT.STD.3.CONC. | Concentration of internal standard #3 |
| INT.STD.1 | Internal standard #1 |
| INT.STD.2 | Internal standard #2 |
| INT.STD.3 | Internal standard #3 |
| ANAL.LAB | Analytical lab that performed analysis on samples |
| USER.TAPE.NO | Tape number assigned by analytical laboratory |

| | |
|---|---|
| CLEAN.TAPE.NO | Tape number where .CLN file is stored |
| MIN.MAX.TIC | Minimum and maximum acceptable peak height with background removed |
| ANAL.DATE | Date GC/MS run is analyzed by AERL |
| REJ.CODE | Code denoting status of GC/MS run |
| INT.STD | Code denoting the technique which located the internal standard |
| IS.TYPE | Code denoting the internal standard |

```
                                            NO. OF     NO. OF   NO. OF
DATABASE NAME                               RECORDS    FIELDS   BYTES
------------------------------              -------    ------   -----
TAPE.LOG                                     15000       36      486


FIELD                       FIELD   FIELD  NO. OF          BYTE
 NO.  FIELD NAME            WIDTH   TYPE   DEC.     SIZE   OFFSET
----- ----------------      -----   -----  ------   ----   ------
  01  RUN.NAME                  6     I       0        4        0
  02  TAPE.NO                   4     C       0        4        4
  03  EXTRACT.NO                6     C       0        6        8
  04  EXTRACT.LOC               8     C       0        8       14
  05  USER.RUN.NAME            26     C       0       26       22
  06  USER.FILE.NO              3     I       0        4       48
  07  USER.SAMPLE.NO           26     C       0       26       52
  08  EPA.SAMPLE.NO            26     C       0       26       78
  09  FRACTION.TYPE             6     C       0        6      104
  10  INDUSTRIAL.CAT           25     C       0       25      110
  11  STAGE.TREATMENT          40     C       0       40      135
  12  DATE.SAMPLED              8     D       0        4      176
  13  DATE.RUN                  8     D       0        4      180
  14  SAMPLER                  50     C       0       50      184
  15  MS                       28     C       0       28      234
  16  GC                       28     C       0       28      262
  17  DATA.SYSTEM              28     C       0       28      290
  18  GC.COLUMN                28     C       0       28      318
  19  AERL.TAPE.NO              4     C       0        4      346
  20  AERL.FILE.NO              3     I       0        4      350
  21  CONC.FACTOR               4     C       0        4      354
  22  INJ.VOLUME                1     C       0        1      358
  23  INT.STD.1.CONC            3     C       0        2      359
  24  INT.STD.2.CONC            3     C       0        2      361
  35  INT.STD.3.CONC            2     C       0        2      363
  26  INT.STD.1                20     C       0       20      365
  27  INT.STD.2                20     C       0       20      385
  28  INT.STD.3                20     C       0       20      405
  29  ANAL.LAB                 25     C       0       25      425
  30  USER.TAPE.NO             20     C       0       20      450
  31  CLEAN.TAPE.NO             2     I       0        4      470
  32  MIN.MAX.TIC               4     I       0        4      474
  33  ANAL.DATE                 8     D       0        4      478
  34  REJ.CODE                  2     C       0        2      482
  35  INT.STD                   1     C       0        1      484
  36  IS.TYPE                   1     C       0        1      485
```

```
C
C
C     DESCRIPTION OF SPECIF.DAT - INPUT CONTROL FILE CREATED BY INFORM
C     (FIVE RECORDS PER SPECTRUM) FOR TAPE CONVERSION
C
C     FIRST INPUT RECORD
C
C     FIELDS 2-5  4-BYTE CHARACTER STRING IDENTIFYING MAG TAPE CONTROLLER
C
C
C     BYTE    DESCRIPTION      VARIABLE
C     ====    ===========      ========
C
C     FIRST SPECTRUM RECORD
C
C     01-06   RUN NAME         RNAME    L*1 (6)
C     07-10   TAPE NUMBER      NUMTAP   I*4
C     11-36   USER RUN NAME    URNAME   L*1 (26)
C     37-39   USER FILE #      UFNUMB   I*4
C     40-42   TIME OFFSET      OFFSET   I*4
C     43-46   AERL TAPE #      AERLTN   I*4
C
C     SECOND SPECTRUM RECORD
C
C     01-06   EXTRACT NUMBER   EXTNUM   I*4
C     07-32   EPA SAMPLE #     EPANUM   L*1 (26)
C
C     THIRD SPECTRUM RECORD
C
C     01-06   FRACTION #       FRACT    L*1 (6)
C     07-31   INDUSTRIAL CODE  INDCOD   L*1 (25)
C     32-71   TREATMENT STAGE  TRTMNT   L*1 (40)
C
C     FOURTH SPECTRUM RECORD
C
C     01-28   MS               MS       L*1 (28)
C     29-56   GC COLUMN        GCCOL    L*1 (28)
C
C     FIFTH SPECTRUM RECORD
C
C     01-25   ANAL LAB         ANALAB   L*1 (25)
C
C     ================================
C
C
```

```
C
C
C
C       DESCRIPTION OF ASSIGNMENT FOR LOGICAL UNIT NUMBERS FOR TAPE CONVERSION
C
C       ======================================================================
C
C       LUN      VAR    DESCRIPTION
C       ===      ===    ===========
C
C        1      INTAPE  TAPE INPUT UNIT
C
C        2      OUTAPE  TAPE OUTPUT UNIT
C
C        3      ISPEC   INPUT FILE FOR DESCRIPTION OF INPUT
C                       TAPE FILES ([200,202]SPECIF.DAT)
C        4      IPLX    OUTPUT PLOT FILE (NAME OF FILE IS OBTAINED
C                       FROM SPECIF.DAT - RUN_NAME.PLX)
C        5      INOUT   LUN FOR INTERACTIVE USE
C
C        6      IMSG    MESSAGE FILE (LINE PRINTER LISTING)
C
C        7      ITMP    TEMPORARY STORAGE FILE
C
C        8      INFORM  OUTPUT FILE FOR UPDATE OF INFORM DATA BASE
C                       (CONTAINS INFO. FOR UPDATE OF AERL TAPE & FILE #)
C                       ( DB1:AERL.DAT )
C        9      IBATCH  OUTPUT BATCH FILE FOR PLOTTING THE PLOTS FOR EACH RUN
C                       ( DB1:PLOT.BIS )
C
C
C       ======================================================================
C
C
```

## DOCUMENTION ON FORTRAN CALLABLE QIO (USED IN TAPE CONVERSION)

```
CALL WTQIO(IFCT,LUN,EVENT,PRI,ISTAT,IPRM,ISW)
```

```
IFCT  - INTEGER WORD   I/O FUNCTION CODE
LUN   - INTEGER WORD   LOGICAL UNIT NUMBER
EVENT - INTEGER WORD   EVENT FLAG
PRI   - INTEGER WORD   QIO REQUEST PRIORITY
ISTAT - 2 INTEG. WORD  I/O STATUS BUFFER
        WORD 1 - BYTE 0: I/O STATUS CODE
                 BYTE 1: 0  (NORMALLY UNUSED)
        WORD 2 - FOR TRANSFER REQUESTS: TOTAL # BYTES
                                        INVOLVED IN TRANSFER
                 FOR OTHER REQUESTS: RETURN OF OTHER STATUS INFO.
IPRM  - 6 INTEG. WORD  DEVICE DEPENDENT PARAMETERS
ISW   - INTEGER WORD   DIRECTIVE STATUS
```

( SEE IAS/RSX-11D FORTRAN SPECIAL SUBROUTINES REFERENCE MANUAL. CHAPTER 5 SECTION 11 )

### STATUS CODES:

RETURNED IN STAT(1)

| CODE | ERROR |
|---|---|
| 1 | SUCCESS |
| 194 | END OF VOLUME |
| 200 | BAD BLOCK |
| 246 | END OF FILE |
| 252 | PARITY ERROR |

( SEE IAS/RSX-11D DEVICE HANDLERS REFERENCE MANUAL, APPENDIX A, PG. 2 FOR A MORE COMPLETE LISTION OF ERROR RETURNS. TO CONVERT TO DEC SYSTEM STANDARD CODES IF THE VALUE RETURNED IN ISTAT(2) IS GREATER THAN ONE, ADD A -256. TO THE CODE RETURNED IN ISTAT(2).)

| FUNCTION | REQUIRED PARAMETERS | RETURNS |
|---|---|---|
| READ | IFCT=512<br>IPRM(1)=ADDRESS OF BUFFER<br>IPRM(2)=BUFFER SIZE | ISTAT(1)=STATUS<br>ISTAT(2)=# BYTES TRANSF. |
| WRITE | IFCT=256<br>IPRM(1)=ADDRESS OF BUFFER<br>IPRM(2)=# OF BYTES TO TRANSFER | ISTAT(1)=STATUS<br>ISTAT(2)=# BYTES TRANSF. |
| WRITE EOF | IFCT=1536 | ISTAT(1)=STATUS |
| REWIND | IFCT=1280 | ISTAT(1)=STATUS |
| SKIP FILE(S) | IFCT=1312<br>IPRM(1)=# FILES TO SKIP | ISTAT(1)=STATUS<br>ISTAT(2)=ACTUAL # FILES SKIPPED |
| SKIP BLOCK(S) | IFCT=1296<br>IPRM(1)=#BLOCKS TO SKIP | ISTAT(1)=STATUS<br>ISTAT(2)=ACTUAL # BLOCKS SKIPPED |
| SENSE CHARACTERISTICS | IFCT=1360 | ISTAT(2)=CHARACTERISTICS STATUS WORD<br>(SEE IAS/RSX-11D DEVICE HANDLER PAGE 7.7 FOR BIT PATTERN) |
| SET CHARACTERISTICS | IFCT=1344 | |

NOTE: FOR A MORE COMPLETE DESCRIPTION OF FUNCTIONS AVAILABLE FOR MAGNETIC TAPE HANDLER, SEE IAS/RSX-11D DEVICE HANDLERS REFERENCE MANUAL, CHAPTER 7.

```
C       DESCRIPTION OF FILES ON CONVERTED TAPE  (INTERNAL TAPE FORMAT)
C
C
C       BYTE    DESCRIPTION     VARIABLE        SOURCE
C       ====    ===========     ========        ======
C
C       =================== HEADER RECORDS ================================
C
C       FIRST HEADER RECORD
C
C       01-06   RUN NAME        RNAME           FROM INPUT FILE, SPECIF.DAT
C       07-70   RUN CONDITIONS  ID              FROM INPUT TAPE
C
C       SECOND HEADER RECORD
C
C       01-26   USER RUN NAME   URNAME          FROM INPUT FILE, SPECIF.DAT
C       27-30   NUMBER OF SCANS NRUNS           FROM INPUT TAPE
C               IN RUN
C       31-38   MAXIMUM BASE    MAXPCK          FROM INPUT TAPE
C               PEAK FOUND IN
C               RUN
C
C       THIRD HEADER RECORD
C
C       01-06   TAPE NUMBER     NUMTAP          FROM INPUT FILE, SPECIF.DAT
C       07-09   STARTING MASS   STMAS           FROM INPUT TAPE
C       10-12   HIGH MASS       GMAX            COMPUTED FROM INPUT TAPE
C       13-16   TIME/SCAN       SECSCN          FROM INPUT TAPE
C               (HUNDREDTHS OF A SEC.)
C       17-19   INTEG. TIME     INTIME          FROM INPUT TAPE
C               (MSEC.)
C       20-24   TOTAL RUN TIME  TOTIME          COMPUTED FROM INPUT TAPE
C               (SEC.)
C       25-27   TIME OFFSET     OFFSET          S150 ONLY, FROM WALT
C               (SEC.)
C       28-30   # MASSES/SCAN   NOMASC          FROM INPUT TAPE
C       31-40   TOTAL RUN IC    NORM            FROM INPUT TAPE
C       41-46   EXTRACT NUMBER  EXTNUM          FROM INPUT FILE, SPECIF.DAT
C       47-72   EPA SAMPLE #    EPANUM          FROM INPUT FILE, SPECIF.DAT
C       73-78   FRACTION #      FRACT           FROM INPUT FILE, SPECIF.DAT
C
C       FOURTH HEADER RECORD
C
C       01-40   TREATMENT STAGE TRTMNT          FROM INPUT FILE, SPECIF.DAT
C       41-68   MS              MS              FROM INPUT FILE, SPECIF.DAT
C
C       FIFTH HEADER RECORD
C
C       01-28   GC COLUMN       GCCOL           FROM INPUT FILE, SPECIF.DAT
C       29-53   ANAL LAB        ANALAB          FROM INPUT FILE, SPECIF.DAT
C       54-78   INDUSTRIAL CODE INDCOD          FROM INPUT FILE, SPECIF.DAT
C
C       ===================  SPECTRA RECORDS  ===============================
C
C       FOR EACH SPECTRA, THERE WILL BE A HEADER RECORD FOLLOWED
C       BY MASS/INTENSITY RECORDS. IF THERE ARE LESS THAN 9 MASS/
C       INTENSTITY PAIRS, ONLY THE HEADER RECORD WILL EXIST. A MASS
C       OF ZERO (0) INDICATES THE END-OF-SCAN.
C
```

```
C
C
C       BYTE    DESCRIPTION     VARIABLE        SOURCE
C       ====    ===========     ========        ======
C
C       =============== HEADER RECORD FOR SPECTRA ==========================
C
C       01-08   BASE PEAK       BASEPK          FROM INPUT TAPE
C       09-17   TIC             TNII            FROM INPUT TAPE
C       18-22   ELAPSED TIME    SCANTM          FROM INPUT TAPE
C               FROM BEGINNING
C               OF RUN (SEC.)
C       25-78   MASS/INT PAIRS  ========        UP TO 9 MASS/INTENSITY PAIRS
C
C       IF THERE ARE ADDITIONAL MASS/INTENSITY PAIRS, THEY WILL BE
C       CONTINUED ON ADDITIOAL RECORDS OF THE FOLLOWING FORM:
C
C       01-78   MASS/INT PAIRS  ========        UP TO 13 MASS/INT PAIRS
C
C
C       NOTE:
C
C               MASS/INTENSITY PAIRS ARE WRITTEN IN 2(I3) FORMAT WHERE
C               THE MASSES HAVE A RANGE OF 1-999 AND THE INTENSITY ARE
C               PERCENTAGES MULTIPLIED BY 10 TO OBTAIN INTEGER EQUIVALENTS.
C               THE PERCENTAGES HAVE A RANGE OF .1% - 99.9% WHICH ARE
C               REPRESENTED BY 001-999.  THE 100% PEAK IS REPRESENTED BY
C               999.
C
C
```

## DESCRIPTION OF PLOT BATCH FILE TO PLOT RGC'S

EACH BATCH JOB BEGINS WITH A STANDARD PDP JOB CARD AND ENDS WITH A STANDARD END OF JOB CARD. THE JOB CARD SPECIFIES THE ACCOUNT FROM WHICH THE COMMANDS ARE TO BE EXECUTED. 'PLOT' IS A NON-STANDARD PDS COMMAND.

TO EXECUTE THE COMMANDS.

```
RECORD 1
------ -
$JOB EGD PLOT 9999

RECORD 2
------ -

COMMAND FILE FACTOR
------- ---- ------

$PLOT   801  1.0

    .
    .
    .

LAST RECORD
---- ------

$EOJ
```

NOTE: THE BATCH JOB FILE PROVIDES THE MEANS OF PLOTTING ALL RGC'S THAT ARE GENERATED FROM SUCCESSFULLY COMPLETED CONVERSIONS.

## DESCRIPTION OF AERL UPDATE COMMAND FILE

THIS COMMAND FILE IS BUILT BY THE CONVERSION PROGRAMS TO UPDATE THE AERL.TAPE.NO AND AERL.FILE.NO FIELDS IN THE INFORM DATABASE. THERE IS ONE RECORD PER SUCCESSFULLY 'CONVERTED' RUN ENTERED INTO THE FILE.

| BYTE | DESCRIPTION |
|---|---|
| 01-06 | RUN NAME |
| 07-10 | AERL FILE NUMBER |
| 11-14 | AERL TAPE NUMBER |

# INCOS TAPE FORMAT

```
C       THERE IS A 4 LINE (80 CHARACTERS/LINE) FILE HEADER.
C               THE FIRST LINES IN A VALID .EP FILE WILL BE A FILE
C               HEADER.  FILE HEADERS IN THE MIDDLE OF A .EP FILE RESULT
C               FROM THE USE OF THE EPA/A COMMAND.  THE FILE HEADER IS
C               IDENTIFIABLE BY AN 0 IN THE SCAN NUMBER ENTRY IN THE FIRST
C               LINE.  THE FIRST FILE HEADER LINE CAN BE READ WITH THE SAME
C               FORMAT AS THE FIRST LINE OF A SCAN HEADER.


C       THE FIRST HEADER LINE CAN BE READ AS FOLLOWS.
        INTEGER NAME (6), IDATE (4)      ;  (ASSUMING 2 CHARACTERS/WORD)
        READ (DKS,110) NAME, ISCAN, IDATE, IHOUR, IMIN
110     FORMAT (6A2,1X,I5,2X,4A2,IX,I2)
C       DATA READ:
C       NAME    ;12 CHARACTER NAME OF ORIGINAL DATA FILE
C       ISCAN   ;0 TO FLAG FILE HEADER FIRST LINE
C       IDATE   ;8 CHARACTER DATE AS "MM/DD/YY"
C       IHOUR   ;HOUR AT START OF RUN
C       IMIN    ;MINUTE AT START OF RUN


C       THE SECOND HEADER LINE CAN BE READ AS FOLLOWS.
        INTEGER ISAMP(32),INST(3)
        READ(DSK,120) ISAMP,INST
120     FORMAT(32A2,10X,3A2)
C       DATA READ:
C       ISAMP   ;64 CHARACTER SAMPLE IDENTIFICATION
C       INST    ;6 CHCARACTER INSTRUMENT NAME


C       THE THIRD HEADER LINE CAN BE READ AS FOLLOWS.
        INTEGER ICOND(32)
        READ(DSK,130)  ICOND, SECPSCAN
130     FORMAT(32A2,10X,F6,2)
C       DATA READ;
C       ICOND               ;64 CHARACTER RUN CONDITIONS
C       SECPSCAN            ;SECONDS PER SCAN FOR DATA IN THIS FILE


C       THE FOURTH HEADER LINE CAN BE READ AS FOLLOWS.
        INTEGER ANALYST(4),SUBMITTED(4),ACCOUNT(4),FORMULE(10),LOWMASS,HIMASS
        READ(DSK,140) ANALYST,SUBMITTED,ACCOUNT,FORMULA,LOWMASS,HIMASS
140     FORMAT(6X,4A2,6X,4A2,6X,10A2,5X,I3,IX,I3)
C       DATA READ
C       ANALYST             ;8 CHARACTER 'ANALYST'
C       SUBMITTED           ;8 CHARACTER 'SUBMITTED BY'
C       ACCOUNT             ;8 CHARACTER 'ACCOUNT NO.'
C       FORMULA             ;20 CHARACTER 'FORMULA'
C       LOWMASS             ;LOWEST MASS POSSIBLY RECORDED
C       HIMASS              ;HIGHEST MASS POSSIBLY RECORDED
```

```
C       FOLLOWING A FILE HEADER WILL BE VARIABLE LENGTH BLOCK REPRESENTING
C       ONE SCAN OF DATE (UNTIL THE END OF FILE OR ANOTHER FILE HEADER.)
C       EACH SCAN HAS THE FOLLOWING FORMAT:
C               HEADER LINE, 80 CHARACTERS
C               MASS/INTENSITY PAIRS IN 80 CHARACTER LINES
C               THE LAST LINE WILL HAVE A 0 MASS/INTENSITY AND SPACE FILL


C       THE SCAN HEADER LINE CAN BE READ AS FOLLOWS:
        INTEGER NAME(6),IDATE(4)
        READ(DSK,200) NAME,ISCAN,IDATE,IHOUR,IMIN,JMIN,JSEC,IBASE,BAREA,RIC
200     FORMAT(6A2,IX,I5,2X,4A2,I3,IX,I2,5X,I3,IX,I2,6X,I4,1X,F9.0,5X,F10.0)
C       DATE READ:
C       NAME    ;12 CHARACTER NAME OF ORIGINAL DATE FILE
C       ISCAN   ;SCAN NUMBER OF ORIGINAL DATA
C       IDATE   ;8 CHARACTER DATE AS 'MM/DD/YY'
C       IHOUR   ;HOUR AT START OF RUN


C       IMIN    ;MINUTE AT START OF RUN
C               ;RETENTION TIME AT END OF SCAN FROM START OF RUN
C       JMIN    ;MINUTES
C       JSEC    ;SECS
C       IBASE   ;NOMINAL MASS OF LARGEST PEAK
C       BAREA   ;INTENSITY OF LARGEST PEAK (BEFORE NORMALIZATION)
C       RIC     ;RECONSTRUCTED ION CURRECT = TOTAL INTENSITY RECORDED
C               ;IN THE SCAN (BEFORE NORMALIZATION)


C       THE MASS/INTENSITY LIST CAN BE READ AS FOLLOWS:
        J12=J+12
        READ(DSK,300) (MASS(I),IAREA(I),I=J,J12),LINENUMBER
300     FORMAT(13(I3,I3),I2)
C       DATA READ:
C       MASS(I)         ;NOMINAL MASS (STRICTLY INCREASING)
C                       ;MASS(I) = 0 INDICATES THE END OF THE SCAN
C       IAREA(I)        ;INTENSITY AT MASS(I)
C                       ;SCALED TO INTENSITY OF BASE = 999
C                       ;WILL NEVER BE 0 IF MASS(I) NOT 0
C       LINENUMBER      ;VALID DATA HAS SEQUENTIAL LINENUMBERS
C                       ;STARTING WITH 1 FOR THE FIRST LINE OF MASSES.
C       (THERE WILL BE BLANK FILL AFTER A 0 MASS/INTENSITY.)
C       THERE IS A MAXIMUM OF 999 MASSES AND HENCE A MAXIMUM OF 77 LINES.
```

## HEWLETT PACKARD TAPE FORMAT

EACH FILE BEGINS WITH TWO HEADER RECORDS, BOTH OF WHICH CAN BE IGNORED. THE FIRST CARD IMAGE OF EACH SPECTRUM CONTAINS THE FOLLOWING INFORMATION.

| BYTE LOCATION | FIELD DESCRIPTION |
|---|---|
| 01 - 05 | BLANKS |
| 06 - 10 | SPECTRUM #. RIGHT JUSTIFIED. |
| 11 - 17 | RETENTION TIME IN MINUTES. RIGHT JUSTIFIED. |
| 18 | "." |
| 19 - 20 | RETENTION TIME IN HUNDREDS OF A MINUTE. |
| 21 - 30 | ABSOLUTE TOTAL ABUNDANCE (SUM OF ALL COMPONENT IONS). RIGHT JUSTIFIED. |
| 31 - 34 | MASS TIME TEN, INTEGER FORMAT. RIGHT JUSTIFIED. |
| 35 - 40 | ABSOLUTE ION ABUNDANCE. RIGHT JUSTIFIED. |
| 41 - 50 | SECOND MASS/ABUNDANCE PAIR |
| 51 - 60 | THIRD MASS/ABUNDANCE PAIR |
| 61 - 70 | FOURTH MASS/ABUNDANCE PAIR |
| 71 - 80 | ABUNDANCE PAIR |

Subsequent card images contain eight mass/abundance pairs, with the possible exception of the last card image of the spectrum.

# FILE STRUCTURES

**<EXPLIC>.TIM**
OUTPUT FILE FROM FILE RETRIEVAL PROGRAM
SEQUENTIAL FILE

| **RECORD 1** | **FORMAT** | **BYTES** |
|---|---|---|
| Hundredths of a second/scan | I6 | 01-06 |
| Integration time | I6 | 07-12 |
| Offset (SECS) | I6 | 13-18 |
| # of Masses/Scan | I6 | 19-24 |
| Override flag for Int. Std. location | I1 | 25 |
| G.C. Column | 28A1 | 26-53 |
| Analytical Lab. | 25A1 | 54-78 |
| | | |
| **RECORD 2** | | |
| Fraction Type | 6A1 | 01-06 |
| Internal Standard 1 | 20A1 | 07-26 |
| Int. Standard Conc. 1 | 3A1 | 27-29 |
| Int. Standard 2 | 20A1 | 30-49 |
| Int. Standard Conc. 2 | 3A1 | 50-52 |
| Int. Standard 3 | 20A1 | 53-72 |
| Int. Standard Conc. 3 | 3A1 | 73-75 |
| Concentration Factor | 4A1 | 76-79 |
| Injection Volume | A1 | 80 |
| Industrial Cat. (Code) | 2A1 | 81-82 |

**NOTE - The location of the Int. Std.(s) is determined by CLEANUP and PBM. If these two modules do not locate the Int. Std.(s), then the location(s) found by the modules STANDARD and PEAK are used in the calculations of RRT's. The override flag, if set to 1, can be used to allow STANDARD and PEAK Int. Std. location(s) to take priority over CLEANUP and PBM.

## <EXPLIC>.PBM

OUTPUT FILE FROM PROBABILITY BASED MATCHING SYSTEM (PBM)
DIRECT ACCESS
RECORD SIZE=28 DOUBLE WORDS
UNFORMATTED
1-10 RECORDS PER SCAN

| RECORD 1 | VARIABLE TYPE | # BYTES |
|---|---|---|
| Run Name | L*1 | 6 |
| Scan # | I*2 | 2 |
| # of Matches | I*2 | 2 |
| **RECORDS 2-10** | | |
| Record Number | I*4 | 4 |
| Compound Name | L*1 | 72 |
| K Value | I*2 | 2 |
| Delta K Value | I*2 | 2 |
| Mol. Ion Flag | I*2 | 2 |
| % Contamination | I*2 | 2 |
| % Component | R*4 | 4 |
| Molecular Weight | I*2 | 2 |
| Molecular Formula | L*1 | 20 |
| # Flags | I*2 | 2 |
| Cas # | I*4 | 4 |

### <EXPLIC>.CLN
OUTPUT FILE FROM CLEAN-UP
SEQUENTIAL FILE

| RECORD 1 | FORMAT | BYTES |
|---|---|---|
| (UNUSED) | 1X | 01 |
| Run Name | 6A1 | 02-07 |

For each scan, there will be a header record followed by N mass/int. pair records. A mass of zero indicates the end of a scan. There can be any number of spectral records.

| HEADER RECORD | | |
|---|---|---|
| Scan # | I4 | 01-04 |
| Areal tic | E14.7 | 05-18 |

| MASS/INT. PAIR RECORDS | | |
|---|---|---|
| Mass | I3 | 01-03 |
| Intensity | I3 | 04-06 |
| (Up to 13 pairs per record) | | |
| Mass | I3 | 73-75 |
| Intensity | I3 | 76-78 |

## <EXPLIC>.SEK

OUTPUT FILE FROM TIMSEK
DIRECT ACCESS
RECORD SIZE=29 DOUBLE WORDS
UNFORMATTED
VARIABLE NUMBER OF SCANS
VARIABLE NUMBER OF RECORDS/SCAN

| HEADER RECORD | | VARIABLE TYPE | # BYTES |
|---|---|---|---|
| Maximum Number of Scans | | I*2 | 2 |
| Number Peak delivered by Cleanup | | I*2 | 2 |
| Fraction Type | | L*1 | 6 |

| RECORD 1 | | | |
|---|---|---|---|
| Scan # | | I*2 | 2 |
| # of RRT's and Rel. Con. (# of Int. Std.'s) | | I*2 | 2 |
| # of PBM Matches | | I*2 | 2 |

| RECORD 2 | | | |
|---|---|---|---|
| RRT | 1 TO 10 | R*4 | 4 |
| Internal Standard | RRT's | I*2 | 2 |

| RECORD 3 | | VARIABLE TYPE | # BYTES |
|---|---|---|---|
| Industrial Cat. Code | | I*2 | 2 |
| GC Col. Code | | I*2 | 2 |
| Anal.lab Code | | I*2 | 2 |
| # of Mass/Int. Pairs | | I*2 | 2 |
| Rel. Conc. | | R*4 | 4 |
| . | 1 TO 10 | | |
| . | | | |
| . | REL. CONC.'S | | |
| Rel. Conc. | | R*4 | 4 |

### PBM MATCH RECORDS (0 TO 5 RECORDS)

| | | |
|---|---|---|
| Record Number | I*4 | 4 |
| Compound Name | L*1 | 72 |
| K Value | I*2 | 2 |
| Delta K Value | I*2 | 2 |
| Mol. Ion Flag | I*2 | 2 |
| % Contamination | I*2 | 2 |
| % Component | R*4 | 4 |
| Molecular Weight | I*2 | 2 |
| Molecular Formula | L*1 | 20 |
| # Flag | I*2 | 2 |
| CAS # | I*4 | 4 |

### MASS/INT. PAIR RECORDS (VARIABLE NUMBER)

| | | | |
|---|---|---|---|
| Mass | | I*2 | 2 |
| Intensity | 13 pairs | I*2 | 2 |
| . | | | |
| . | per | | |
| . | | | |
| Mass | record | I*2 | 2 |
| Intensity | | I*2 | 2 |

## HIT RRT FILE (HISLIB.RRT)

DIRECT ACCESS
UNFORMATTED
RECORD SIZE=450 DOUBLE WORDS

POINTER TO LAST RECORD IN FILE

| **RECORD 1** | **VARIABLE TYPE** | **BYTE LOCATION** |
|---|---|---|
| Record # | I*4 | 01-04 |
| Entry # | I*2 | 05-06 |

| **RECORD 2 TO N** | (1800 BYTE RECORDS) | |
|---|---|---|
| Rel. Retention Time | R*4 | 01-400 |
| Int. Std. Code | I*2 | 401-600 |
| GC Col. Code | I*2 | 601-800 |
| Anal. Lab. Code | I*2 | 801-1000 |
| CAS # | I*4 | 1001-1400 |
| Pointer to entry in HISLIB.HIT | I*4 | 1401-1800 |

18 Bytes/Entry
100 Entries/Record

## FORTRAN READ FOR FILE

```
      REAL*4 RRT(100)
      INTEGER*4 CAS(100),POINT(100),RECORD
      INTEGER*2 INTSTD(100),GC(100),ANALAB(100)ENTRY
      OPEN(UNIT=1,NAME='HISLIB.RRT',ACCESS='DIRECT',TYPE='OLD')
C     READ LENGTH OF FILE
      READ(1'1)RECORD,ENTRY
      IREC=2
      READ(1'IREC)RRT,INTSTD,GC,ANALAB,CAS,POINT
```

## MATCHED SPECTRA FILE (HISLIB.HIT)

DIRECT ACCESS
UNFORMATTED
RECORD SIZE=28 DOUBLE WORDS

### RECORD 1

| | |
|---|---|
| Next available record in file | I*4 |

### N NUMBER OF HIT RECORDS (3 RECORDS/HIT)

### HIT RECORD 1

| | | |
|---|---|---|
| Run Name | L*1 | 6 |
| Scan # | I*2 | 2 |
| Industry Code | I*2 | 2 |
| GC Col. Code | I*2 | 2 |
| Anal. Lab Code | I*2 | 2 |
| # of Int. Std.'s | I*2 | 2 |
| Record # pointer to RRT.HIT file | I*4 | 4 |
| Entry # pointer to RRT.HIT file | I*4 | 4 |
| Flag 1 for standard runs | L*1 | 1 |
| Flag 2 for who identified compound | L*1 | 1 |
| Date of identification | L*1 | 9 |
| Fraction Type | L*1 | 6 |
| Point to next compound matched with same RRT window | I*4 | 4 |
| Total # of peaks found by cleanup | I*2 | 2 |
| Flag 3 for blank runs | L*1 | 1 |
| Flag 4 for spiked runs | L*1 | 1 |

***NOTE***

a. The flag for specifying whether the entry came from a standard run or not contains one of the following values:

1) FLAG1=0 NON-STANDARD RUN
2) FLAG1=1 STANDARD RUN (LAST THREE CHARACTERS OF FRACTION TYPE CONTAIN 'STD')

b. The flag for specifying who identified compound contains one of the following values:

1) FLAG2=0 HISLIB IDENTIFIED
2) FLAG2=1 CHEMIST IDENTIFIED (NEW COMPOUND)
3) FLAG2=2 CHEMIST IDENTIFIED (RRT OUTSIDE OF WINDOW)
4) FLAG2=3 CHEMIST IDENTIFIED (NO PBM MATCH AVAILABLE)

c. The flag for determining if this run was a blank or not contains one of the following values:

1) FLAG3=0 NON-BLANK RUN
2) FLAG3=1 BLANK RUN

d. The flag for determining if this run was a spiked run or not contains one of the following values:

1) FLAG4=0 NON-SPIKED RUN
2) FLAG4=1 SPIKED RUN

**HIT RECORD 2**

| | | | |
|---|---|---|---|
| Int. Std. Code | \ | I*2 | 2 |
| RRT | 1 to 10 sets | R*4 | 4 |
| Rel. Conc. | / | R*4 | 4 |

**HIT RECORD 3**

| | | |
|---|---|---|
| CMP Name | L*1 | 72 |
| K Value | I*2 | 2 |
| Delta K Value | I*2 | 2 |
| Mol. Ion Flag | I*2 | 2 |
| % Contamination | I*2 | 2 |
| % Component | R*4 | 4 |
| Molecular Weight | I*2 | 2 |
| Molecular Formula | L*1 | 20 |
| # Flags | I*2 | 2 |
| CAS # | I*4 | 4 |

**<EXPLIC>.HLD**

OUTPUT FILE FROM TIMSEK
DIRECT ACCESS
RECORD SIZE=29 DOUBLE WORDS
UNFORMATTED
VARIABLE NUMBER OF SCANS
VARIABLE NUMBER OF RECORDS/SCAN

| **HEADER RECORD** | **VARIABLE TYPE** | **# BYTES** |
|---|---|---|
| Maximum Number of Scans | I*2 | 2 |
| Number Peak delivered by Cleanup | I*2 | 2 |
| Fraction Type | L*1 | 6 |

| **RECORD 1** | | |
|---|---|---|
| SCAN # | I*2 | 2 |
| # OF RRT's and Rel. Con.<br>(# of Int. Std.'s) | I*2 | 2 |
| # of PBM Matches | I*2 | 2 |

| **RECORD 2** | | | |
|---|---|---|---|
| RRT | 1 TO 10 | R*4 | 4 |
| Internal Standard | RRT's | I*2 | 2 |

| **RECORD 3** | **VARIABLE TYPE** | **# BYTES** |
|---|---|---|
| Industrial Cat. Code | I*2 | 2 |
| GC Col. Code | I*2 | 2 |
| Anal.lab Code | I*2 | 2 |
| # of Mass/Int. Pairs | I*2 | 2 |
| Rel. Conc. | R*4 | 4 |
| . 1 TO 10 | | |
| . | | |
| . REL. CONC.'S | | |
| Rel. Conc. | R*4 | 4 |

**<EXPLIC>.BCK**

SAME FORMAT AS <EXPLIC>.DAT

### GCCOL.DAT

FILE OF PRESENT GC COLUMNS BEING USED IN THE HISTORICAL DATABASE
SEQUENTIAL FILE
ONE RECORD GC COLUMN
N NUMBER OF RECORDS

| FIELD | VARIABLE TYPE | # BYTES |
|---|---|---|
| GC Column | L*1 | 28 |
| Alternate GC Column code | I*2 | 02 |
| Start time (sec.) for Window for 1st Int. Std. | I*2 | 06 |
| Stop time (sec.) for Window for 1st Int. Std. | I*2 | 06 |
| Start time (sec.) for Window for 2nd Int. Std. | I*2 | 06 |
| Stop time (sec.) for Window for 2nd Int. Std. | I*2 | 06 |

### ANALAB.DAT

FILE OF PRESENT ANALYTICAL LABORATORIES BEING USED WITH THE HISTORICAL DATABASE
SEQUENTIAL FILE
ONE RECORD/ANALYTICAL LAB
N NUMBER OF RECORDS

| FIELD | VARIABLE TYPE | # BYTES |
|---|---|---|
| Analytical Laboratory | L*1 | 25 |

### INTSTD.DAT

FILE OF PRESENT INTERNAL STANDARDS BEING USED IN THE HISTORICAL DATABASE
SEQUENTIAL FILE
ONE RECORD PER INTERNAL STANDARD
N NUMBER OF RECORDS

| FIELD | VARIABLE TYPE | # BYTE |
|---|---|---|
| Internal Standard | L*1 | 20 |

## <RUN NAME>.STD

SEQUENTIAL FILE
FILE CONTAINS THE LOCATION AND AREA UNDER CURVE OF BASE PEAK FOR THE INTERNAL STANDARD(S).
N NUMBER OF RECORDS (DEPENDENT UPON NUMBER OF INT. STD.)

| RECORD<br>FIELD | VARIABLE TYPE | # BYTES |
|---|---|---|
| Scan # | I*2(I4) | 4 |
| Area of Peak | R*4(F10.0) | 10 |

## MASS.TMP

TEMPORARY FILE OF MASSES FOR INTERNAL STANDARD(S).
SEQUENTIAL FILE

| RECORD 1<br>FIELD | VARIABLE TYPE | # BYTES |
|---|---|---|
| Number of Internal Standards | I*2 | 2 |
| **RECORD 2** | | |
| Number of Masses for 1st Int. Std. | I*2 | 2 |
| **RECORD 3** | | |
| Mass for 1st Int. Std. | I*3 | 3 |
| . | . | . |
| . | . | . |
| . | . | . |
| Mass for 1st Int. Std. | I*3 | 3 |
| **RECORD 4 (optional)** | | |
| Number of Masses for 2nd Int. Std. | I*2 | 2 |
| **RECORD 5 (optional)** | | |
| Mass for 3nd Int. Std. | I*2 | 3 |
| . | . | . |
| . | . | . |
| . | . | . |
| Mass for 3nd Int. Std. | I*2 | 3 |

**MASS_STD**

FILE OF MASSES FOR THE INTERNAL STANDARDS
(D10 ANTHRACENE, BROMOCHLOROMETHANE,
AND 1,4-DICHLOROBUTANE)

**RECORD 1**

| **FIELD** | **VARIABLE TYPE** | **# BYTES** |
|---|---|---|
| Dollar Sign<br>(signfies start of new Int. Std.) | L*1 | 1 |
| Compound Name | L*1 | 20 |
| **RECORD 2** | | |
| Number of Masses | I*2 | 3 |
| **RECORD 3** | | |
| Mass | I*2 | 3 |
| . \ | . | . |
| . 1-15 masses | . | . |
| . / | . | . |
| Mass | I*2 | 3 |
| **RECORD 4 (OPTIONAL)** | | |
| Mass | I*2 | 3 |
| . \ | . | . |
| . 1-15 masses | . | . |
| . / | . | . |
| Mass | I*2 | 3 |

***NOTE - Presently this file contains the mass/intensity pairs for 4 internal standards:

| INTERNAL **STANDARD** | **# MASSES** | **MASSES** |
|---|---|---|
| 1. BROMOCHLOROMETHANE | 8 | 49,130,128,51<br>93,132,95,47 |
| 2. 1,4-DICHLOROBUTANE | 7 | 55,63,90,62<br>54,92,64 |
| 3. D10 ANTHRACENE | 8 | 188,189,187,186<br>184,160,94,80 |
| 4. 2BROMO1CHLOROPROPANE | 8 | 77,76,79,49,63,<br>107,109,78 |

## GCWINDOW.MIS

DIRECT ACCESS
UNFORMATTED
180 BYTE/RECORD
1 RECORD/UNIQUE GC COLUMN

| BYTE LOCATION | FIELD DESCRIPTOR | VARIABLE TYPE |
|---|---|---|
| **RECORD 1** | | |
| 01-04 | Pointer to next available record in file | I*4 |
| **RECORDS 2 THRU N** | | |
| 01-04 | Pointer to records in the research file (RRT.MIS) that contains the RRT(s) from .0 to .1 for this GC Column | I*4 |
| 05-08 | ... RRT(s) from .1 to .2 for this GC Column | I*4 |
| 09-12 | ... RRT(s) from .2 to .3 for this GC Column | I*4 |
| | .<br>.<br>. | |
| 177-180 | ... RRT(s) from 4.4 to 4.5 for this GC Column | I*4 |

***NOTE - N REPRESENTS THE NUMBER OF GC Column presently in use plus one.

I.E. - (NUMGC + 1) = N

**FORTRAN READ FOR FILE:**

```
        INTEGER*4 NUMREC,POINT(45)
        OPEN(UNIT=1,NAME='GCWINDOW.MIS',TYPE='OLD',ACCESS='DIRECT')
C       READ LENGTH OF FILE
        READ(1'1)NUMREC
C       SET UP FOR READ
        IREC=GCCODE+1
        READ(1'IREC)POINT
```

***NOTE***

1. The GC Column code found in the <Runname>.MIS file corresponds to the (GCCODE + 1) record in this file.

   (GCCODE = GC COLUMN CODE)

2. A pointer value of zero represents no entries within that RRT range.

**RRT.MIS (RRESEARCH FILE)**
DIRECT ACCESS
UNFORMATTED
1412 BYTES/RECORD

| **BYTE LOCATION** | **FIELD DESCRIPTOR** | **VARIABLE TYPE** |
|---|---|---|
| **RECORD 1** | | |
| 01-04 | Pointer to next available record in file | I*4 |
| **RECORDS 2 THRU LAST** | | |
| 01-02 | Last entry used in record | I*2 |
| 03-06 | Record pointer in file GCWINDOW.MIS | I*4 |
| 07-08 | Entry pointer in file GCWINDOW.MIS | I*2 |
| 09-408 | RRT | R*4 |
| 409-808 | 'Chaining pointer' to spectrum | I*4 |
| 809-1008 | Mass of 100% peak | I*2 |
| 1009-1208 | Int. Std. code | I*2 |
| 1209-1408 | Analytical lab code | I*2 |
| 1409-1412 | 'Chaining pointer' to next record containing RRT's that 'fall within' the window described by GCWINDOW.MIS | I*4 |

**FORTRAN READ OF FILE:**

```
        REAL*4 RRT(100)
        INTEGER*4 POINT(100), RECNUM,NXTREC,LAST
        INTEGER*2 MAS100(100),INTSTD(100),ANALAB(100)
        INTEGER*2 ENTNUM,LSTENT

        OPEN(UNIT=1,TYPE='OLD',NAME='RRT.MIS',ACCESS='DIRECT')

        IREC=1
        READ(1'IREC)LAST

        IREC=2
        READ(1'IREC)LSTENT,RECNUM,ENTNUM,RRT,POINT
                    MAS100,INTSTD,ANALAB,NXTREC
```

***NOTE - If the 'chaining pointer' to next record containing RRT's is zero, then no further records are chained to this RRT window.

**HISLIB.MIS**
DIRECT ACCESS
UNFORMATTED
2 RECORDS/SPECTRUM ENTRY
22 BYTE/RECORD

| BYTE LOCATION | FIELD DESCRIPTOR | VARIABLE TYPE |
|---|---|---|
| **RECORD 1 (OF SPECTRUM)** | | |
| 01-04 | RRT | R*4 |
| 05-08 | 'Chaining' pointer | I*4 |
| 09-12 | Record pointer to RRT.MIS | I*4 |
| 13-14 | Entry pointer | I*2 |
| 15-16 | Scan # | I*2 |
| 17-22 | Run name | L*1 |
| 23-24 | # scans retrieved by CLEANUP | I*4 |
| **RECORD 2 (OF SPECTRUM)** | | |
| 01-02 | MASS/INT. PAIR | I*2 |
| . | . | . |
| . | . (8 PAIRS) | . |
| . | . | . |
| 15-16 | MASS/INT. PAIR | I*2 |
| 17-20 | CONCENTRATION | R*4 |
| 21-24 | DATE (JULIAN AND YEAR) | I*4 |

***NOTE***

Mass/intensity pairs are stored in a integer word (2 bytes - 16 bits) where the low order 7 bits represent the intensity and the remaining 9 bits are used to store the mass value.

Mass range: 1-511 mass units
Intensity range: 1-100 1% intensity units

```
C
C                         LIST.MAS FILE FORMAT
C
C
C         A .MAS file (referred to as a master file) consists of a header
C     record followed by several data records. The logical record length
C     of the file is 1024 words. The formats of the header record and of the
C     data records are described below.
C         Any additions, deletions, alterations, etc. that are made to
C     the layout of the header record or data record should be indicated
C     in this description.
C
C
C
C                         HEADER RECORD FORMAT
C
C
C                              +-----------------------+
C LIBMAP        1              |       bitmap of       |
C                              |       .MAS file       |
C                              |           .           |
C                              |           .           |
C               64             |           .           |
C                              +-----------------------+
C LIBSIZ        65             |    # data records     |         1
C                              +-----------------------+
C CONFLG        66             |    conversion flag    |
C                              +-----------------------+
C               67             |     high- or low-     |
C                              |      resolution       |
C                              |       specific        |
C                              |      information      |
C               99             |           .           |
C                              +-----------------------+
C SCHDEV        100            |    search library     |
C               104            |        device         |
C                              +-----------------------+
C SCHLIB        105            |    search library     |
C               109            |         name          |
C                              +-----------------------+
C EXPDEV        110            |      experiment       |
C               114            |        device         |
C                              +-----------------------+
C EXPNAM        115            |      experiment       |
C               120            |         name          |
C                              +-----------------------+
C STDLIB        121            | calibration standards |
C               124            |        library        |
C                              +-----------------------+
C STDDEV        125            | calibration standards |
C               129            |        device         |
C                              +-----------------------+
C UNIT          130            |     print unit #      |         6
C                              +-----------------------+
C DELFLG        131            |  coptap delete flag   |
C                              +-----------------------+
C
```

```
C
```

```
C
C                          +-----------------------+
C QUECNT        132        |  experiment counter   |
C                          +-----------------------+
C QUEMAX        133        | # experiments queued  |
C                          +-----------------------+
C QUESIZ        134        |  max queue capacity   |
C                          +-----------------------+
C LOCLIB        135        |     local library     |
C               139        |         name          |
C                          +-----------------------+
C LOCDEV        140        |     local library     |
C               144        |        device         |
C                          +-----------------------+
C PLTSET        145        |  plttic overlay flag  |
C                          +-----------------------+
C MOLFLG        146        |  molion overlay flag  |
C                          +-----------------------+
C CLNFLG        147        |  cleanu overlay flag  |          0
C                          +-----------------------+
C STDFLG        148        |  time standards flag  |
C                          +-----------------------+
C TSKFLG        149        |  timsek overlay flag  |
C                          +-----------------------+
C LOCFLG        150        |  srchlc overlay flag  |
C                          +-----------------------+
C SCHFLG        151        |  search overlay flag  |
C                          +-----------------------+
C NULL          152        |  Ret. index search    |
C                          |  window. Def = 20     |
C                          +-----------------------+
C NULL          153        |                       |
C               154        |        unused         |
C                          +-----------------------+
C HSTNAM        155        |     history file      |
C               159        |         name          |
C                          +-----------------------+
C HSTDEV        160        |     history file      |
C               164        |        device         |
C                          +-----------------------+
C IPFLAG        165        |  cleanu print flag    |          1
C                          +-----------------------+
C EXTCNT        166        | coptap extension count|
C                          +-----------------------+
C EXTLST        167        |      extension 1      |
C               168        |                       |
C                          +-----------------------+
C               169        |           .           |
C                          |           .           |
C               184        |           .           |
C                          +-----------------------+
C               185        |      extension 10     |
C               186        |                       |
C                          +-----------------------+
C CHKSUM        187        |    checksum array     |
C               486        |   1 per slot in file  |
C                          +-----------------------+
C NULL          487        |  fixed ret.time flag  |
C                          |  = -1 if FIXRET run   |
C                          +-----------------------+
C               488        |      unused area      |
C                          |           .           |
C              1024        |                       |
C                          +-----------------------+
C
C
C                         DATA RECORD FORMAT
C
C
C                      +-----------------------+
C TOTION        1      |    TIC value this     |
C               2      |       spectrum        |
C                      +-----------------------+
C CSRETX        3      |   cumulative sum of   |
C               4      |   retention indices   |
C                      +-----------------------+
C CSREX2        5      |  cum sum of squared   |
C               6      |   retention indices   |
C                      +-----------------------+
C CSMSCR        7      |     cum sum of max    |
C               8      |     match scores      |
C                      +-----------------------+
C CSMSC2        9      |  cum sum of squared   |
C              10      |  max match scores     |
C                      +-----------------------+
C TOTMAT       11      |   accumulated sum     |
C              12      |   matching scores     |
C                      +-----------------------+
C NULL         13      |       unused          |
C              45      |                       |
C                      +-----------------------+
C IRELCN       46      | rel. con. (integer)   |
C                      +-----------------------+
C RELCON       47      |  relative concent.    |
C              48      |                       |
C                      +-----------------------+
C ACTMAX       49      |  actual maximum amp   |
C              50      |        (r*4)          |
C                      +-----------------------+
C LIBSLT       51      |    library slot #     |
C                      +-----------------------+
C LIBSRC       52      |       library         |
C              56      |        name           |
C                      +-----------------------+
C NITFLG       57      |    nitrogen flag      |
C                      +-----------------------+
C ARFLG        58      |    aromatic flag      |
C                      +-----------------------+
C MOLINT       59      |     molecuar ion      |
C              63      | candidate intensities |
C                      +-----------------------+
C SCORE        64      |    matching score     |
C                      +-----------------------+
C FIGMER       65      |    figure of merit    |
C                      +-----------------------+
C NMATCH       66      |# matches this compound|
C                      +-----------------------+
C MOLCAN       67      |     molecular ion     |
C              71      |      candidates       |
C                      +-----------------------+
C TICRAT       72      |    cleanup ratio      |
C                      +-----------------------+
C MOLWT        73      |   molecular weight    |
C                      +-----------------------+
C SCAN2        74      |   hi mass scanned     |
C                      +-----------------------+
C SCAN1        75      |   lo mass scanned     |
C                      +-----------------------+
C
C                       +-----------------------+
C THRESH        76      | background threshold  |
C                       +-----------------------+
C MAXAMP        77      |   maximum amplitude   |
C                       +-----------------------+
C SPECNO        78      |      spectrum #       |
C                       +-----------------------+
C LDATE         79      |    date entered in    |
C               83      |        .MAS file      |
C                       +-----------------------+
C EXPNAM        84      |       experiment      |
C               88      |         name          |
C                       +-----------------------+
C INTG          89      |    integration time   |
C                       +-----------------------+
C GCTIME        90      |    retention index    |
C                       +-----------------------+
C SOURCE        91      |     source number     |
C                       +-----------------------+
C RDATE         92      |      date spectrum    |
C               96      |         run           |
C                       +-----------------------+
C NAME          97      |        chemical       |
C               128     |         name          |
C                       +-----------------------+
C DENDRA        129     |       DENDRAL         |
C               253     |      structure        |
C                       +-----------------------+
C RELtic          254   |  Tic*H/A              |
C               255     |   H = height,         |
C                       |   A = area  of model. |
C                       +-----------------------+
C AMP           257     |    mass amplitudes    |
C                       |           .           |
C                       |           .           |
C                       |           .           |
C                       |           .           |
C               1024    |           .           |
C                       +-----------------------+
        END
```

## FINNIGAN FILE FORMATS

The finnigan data collection system utilizes three files to store data collected. These files have extensions PAR (experiment parameters), DIR (total ion current information), and DAT (mass amplitudes), respectively.

Described below are the formats of each of these files.

The program READ can be used to read any of these files, which are stored in direct-access format on our system.

## <RUN NAME>.PAR FILE FORMAT

The par file consists of two logical records, each with length 256. The first record contains ASCIZ strings describing experiment name, operator, date, time, and chemical name. The second record contains experiment parameters such as mass range scanned, # masses scanned, integration time, etc; in absolute binary form.

[  *  indicates that the values are determined in the BUILD program  ]
[  a blank to the right of the box indicates value not used in our system]

## Record 1

```
                           +-------------------------+
EXPID            1         |  experiment id          |             *
                           +-------------------------+
OPER             6         |  operator name          |
                           +-------------------------+
PDATE            11        |  date run               |             *
                           +-------------------------+
PTIME            16        |  time run               |             *
                           +-------------------------+
CHEMN            21        |  chemical name          |
                           +-------------------------+
CALIB            59        |  calibration file       |
                 63        |  used. default PA190.   |
                           +-------------------------+
TIMSTD           64        |    time standards       |   12,18,24,bb,bb
                 68        |        array            |
                           +-------------------------+
DEGREE           69        |    temperature rate     |   8 for base/ 10 for acid
                           +-------------------------+
MINSTD           70        |  minimum standard pres  |             12
                           +-------------------------+
MAXSTD           71        |  maximum standard pres  |             24
                           +-------------------------+
NORATE           72        | # allow. prog. rates=N  |             1
                           +-------------------------+
RATES            73        |   programming rates     |    8 for base/ 10 for acid
                 72 + N    |   array. (N leq 6)      |
                           +-------------------------+
TIMSEK WINDOWS   73 + N    | DEFAULT TO              |   output(74)=bb
                 73 + N+4  | 30,12,12,12,12          |
                           +-------------------------+
COMENT           84        |  operator's 2 cents     |
                 108       |     (comment)           |
                           +-------------------------+
WHICHS           109       |  Which standard used    |
                           |  for rel. conc. calc.   |
                           +-------------------------+
                 110       |  Number of masses in    |
                           |  following list.        |
                           +-------------------------+
                 111       |  64 word mass list of   |
                           |  masses in mass frag    |
                           |  run used by CLEANU.    |
                           |  set un by EXpand in    |          :
                 174       |  MASTER, or by PEDIT.   |          ^
                 175       +-------------------------+
                 256       |     UNUSED              |
                           +-------------------------+          Io
                                        *
                                -- 0 --   ^.  Titan
```

## Record 2

```
                            +------------------------+
FORMAT          0           |  format of collection  |
                            +------------------------+
BLKCNT          1           |  # blocks DAT file     |     # of spectra
                            +------------------------+
SPEC            2           |  # spectra collected.  |     # of spectra
                            |  See word 67. This val.|
                            |  is less than or equal |
                            |  to word 67.           |
                            +------------------------+
EXTIME          3           |  time of experiment    |     min and sec
                4           |      (2 words)         |
                            +------------------------+
LOMASS          5           |  lo mass # scanned     |     *
                            +------------------------+
HIMASS          6           |  hi mass # scanned     |     *
                            +------------------------+
POINTS          7           |  # masses per spectra  |     *
                            +------------------------+
TAPCNT          8           |  # DECtapes used       |
                            +------------------------+
MAXBLK          9           | max # blocks necessary |
                            +------------------------+
MAXTAP          10          | max # tapes necessary  |
                            +------------------------+
BLKTAP          11          |  # blocks per tape     |
                            +------------------------+
BLKSIZ          12          |   block size (words)   |
                            +------------------------+
MINTIC          13          |     min TIC value      |
                            +------------------------+
MAXTIC          14          |     max TIC value      |
                            +------------------------+
MASPAR          15          |     # mass pairs       |
                            +------------------------+
RIwind          16          | Retention index window |
                            | for search.(default 20)|
                            +------------------------+
CupMin          17          | Minimum total ion cur- |     1000
                            | rent for cleanup.      |
                            +------------------------+
CupMax          18          | Maximum total ion cur- |     1000
                            | for cleanup spectrum   |
                            | detection.             |
                            +------------------------+
DubPre          19          | =0  if sing. prec.     |     0
                            | =1  if double prec.    |
                            | This is .DAT file info |
                            +------------------------+
Left            20          | Points to left of max. |     2
                            +------------------------+
Right           21          | Points to rigt of max. |     2
                            +------------------------+
IsFrag          22          | Signals frag run for   |     0
                            | CLEANUP                |
                            +------------------------+
                        +------------------------+
TopRma          23      | Top Rate Maximum Mass  |        1000
                        +------------------------+
ScnDir          24      | = 1 if LOW to HIGH mass|        1
                        | =-1 if HIGH to LOW mass|
                        +------------------------+
NULL            25      |     this space         |
                        |       is               |
                        |                        |
                63      |      UNUSED            |
                        +------------------------+
MINTIM          64      | min time per spectrum  |        3
                        +------------------------+
INTEG           65      |   integration time     |        *
                        +------------------------+
PREC            66      | # points scanned/mass  |        1
                        +------------------------+
MAXSP           67      | max # spectra to scan. |        # of spectra
                        | The number of spec. for|
                        | which the .DAT file is |
                        | ALLOCATED.             |
                        +------------------------+
DEVICE          68      |  rad50 collection dev  |
                        +------------------------+
R50EXP          69      |  rad50 experiment id   |
                70      |      (2 words)         |
                        +------------------------+
SATVAL          71      |  Saturated value       |        *
                        +------------------------+
BACKGR          72      |  Background level      |        *
                        |  for RGC.              |
                        |------------------------+
NULL            73      |         unused         |
                        |           .            |
                        |           .            |
                        |           .            |
                127     |           .            |
                        +------------------------+
BITMAP          128     |     bitmap of masses   |
                        |           .            |
                        |           .            |
                        |           .            |
                191     |           .            |
                        +------------------------+
MASLST          192     |  list of mass pairs    |
                        |           .            |
                        |           .            |
                        |           .            |
                255     |           .            |
                        +------------------------+
```

## <RUN NAME>.DIR FILE FORMAT

The DIR file contains a TIC (Total Ion Current) value for each spectrum collected and the elapsed time in seconds since the beginning of the experiment.
The logical record length of the DIR file is 4 words. The number of records in a given DIR file may be found by looking at word 2 of record 2 of the PAR file (# spectra).

```
                              Record m

                        +------------------------+
EXTIME          0       |      elapsed time      |        I*4
                        +------------------------+
TICVAL          2       |       TIC value        |        I*4
                        +------------------------+
```

## <RUN NAME>.DAT FILE FORMAT

The DAT file contains the mass amplitude values collected. There is one record for each spectrum collected (this number is found in word 2 of record 2 of the PAR file). The logical record length is equal to the number of masses scanned in each spectrum (found in word 7 of record 2 of the PAR file). The amplitudes are packed so that a mass which is not scanned has no entry in the record.
The number of masses scanned and the masses scanned remain constant throughout the experiment.

```
                              Record m

                        +------------------------+
SPECTR          0       |    amplitude mass1     |        ILOW
                        +------------------------+
                1       |    amplitude mass2     |
                        +------------------------+
                        |           .            |
                        |           .            |
                        |           .            |
                        +------------------------+
                n-1     |    amplitude massn     |        IHIGH
                        +------------------------+

                N = NMASS
```

```
C       DESCRIPTION OF FILES ON OUTPUT TAPE
C
C
C       BYTE    DESCRIPTION     VARIABLE        SOURCE
C       ====    ===========     ========        ======
C
C       ================== HEADER RECORDS ==================================
C
C       FIRST HEADER RECORD
C
C       01-06   RUN NAME        RNAME           FROM INPUT FILE, SPECIF.DAT
C       07-70   RUN CONDITIONS  ID              FROM INPUT TAPE
C
C       SECOND HEADER RECORD
C
C       01-26   USER RUN NAME   URNAME          FROM INPUT FILE, SPECIF.DAT
C       27-30   NUMBER OF SCANS NRUNS           FROM INPUT TAPE
C               IN RUN
C       31-38   MAXIMUM BASE    MAXPCK          FROM INPUT TAPE
C               PEAK FOUND IN
C               RUN
C
C       THIRD HEADER RECORD
C
C       01-06   TAPE NUMBER     NUMTAP          FROM INPUT FILE, SPECIF.DAT
C       07-09   STARTING MASS   STMAS           FROM INPUT TAPE
C       10-12   HIGH MASS       GMAX            COMPUTED FROM INPUT TAPE
C       13-16   TIME/SCAN       SECSCN          FROM INPUT TAPE
C               (HUNDREDTHS OF A SEC.)
C       17-19   INTEG. TIME     INTIME          FROM INPUT TAPE
C               (MSEC.)
C       20-24   TOTAL RUN TIME  TOTIME          COMPUTED FROM INPUT TAPE
C               (SEC.)
C       25-27   TIME OFFSET     OFFSET          S150 ONLY, FROM WALT
C               (SEC.)
C       28-30   # MASSES/SCAN   NOMASC          FROM INPUT TAPE
C       31-40   TOTAL RUN IC    NORM            FROM INPUT TAPE
C       41-46   EXTRACT NUMBER  EXTNUM          FROM INPUT FILE, SPECIF.DAT
C       47-72   EPA SAMPLE #    EPANUM          FROM INPUT FILE, SPECIF.DAT
C       73-78   FRACTION #      FRACT           FROM INPUT FILE, SPECIF.DAT
C
C       FOURTH HEADER RECORD
C
C       01-40   TREATMENT STAGE TRTMNT          FROM INPUT FILE, SPECIF.DAT
C       41-68   MS              MS              FROM INPUT FILE, SPECIF.DAT
C
C       FIFTH HEADER RECORD
C
C       01-28   GC COLUMN       GCCOL           FROM INPUT FILE, SPECIF.DAT
C       29-53   ANAL LAB        ANALAB          FROM INPUT FILE, SPECIF.DAT
C       54-78   INDUSTRIAL CODE INDCOD          FROM INPUT FILE, SPECIF.DAT
C
C       ==================  SPECTRA RECORDS  ================================
C
C       FOR EACH SPECTRA, THERE WILL BE A HEADER RECORD FOLLOWED
C       BY MASS/INTENSITY RECORDS. IF THERE ARE LESS THAN 9 MASS/
C       INTENSTITY PAIRS, ONLY THE HEADER RECORD WILL EXIST. A MASS
C       OF ZERO (0) INDICATES THE END-OF-SCAN.
C
```

```
C
C
C       BYTE    DESCRIPTION     VARIABLE        SOURCE
C       ====    ===========     ========        ======
C
C       =============== HEADER RECORD FOR SPECTRA =========================
C
C       01-08   BASE PEAK       BASEPK          FROM INPUT TAPE
C       09-17   TIC             TNII            FROM INPUT TAPE
C       18-22   ELAPSED TIME    SCANTM          FROM INPUT TAPE
C               FROM BEGINNING
C               OF RUN (SEC.)
C       25-78   MASS/INT PAIRS  ========        UP TO 9 MASS/INTENSITY PAIRS
C
C       IF THERE ARE ADDITIONAL MASS/INTENSITY PAIRS, THEY WILL BE
C       CONTINUED ON ADDITIOAL RECORDS OF THE FOLLOWING FORM:
C
C       01-78   MASS/INT PAIRS  ========        UP TO 13 MASS/INT PAIRS
C
C
C       NOTE:
C
C               MASS/INTENSITY PAIRS ARE WRITTEN IN 2(I3) FORMAT WHERE
C               THE MASSES HAVE A RANGE OF 1-999 AND THE INTENSITY ARE
C               PERCENTAGES MULTIPLIED BY 10 TO OBTAIN INTEGER EQUIVALENTS.
C               THE PERCENTAGES HAVE A RANGE OF .1% - 99.9% WHICH ARE
C               REPRESENTED BY 001-999.  THE 100% PEAK IS REPRESENTED BY
C               999.
C
C
```

## Description of FOR005.DAT Created by BATCHJOB
## Used During Run Processing

| RECORD | VALUE | DESCRIPTION |
|---|---|---|
| 1 | nnnnn | Run name (left justified, maximum 6 digits) |
| 2 | DB0 | Output destination |
| 3 | Y | CLEANUP printout requested |
| 4 | 1 | ANYTHING ELSE? answer (1=yes) |
| 5 | | DEBUG? answer (no value = no) |
| 6 | 1 | NEW INPUT PARAMETERS? answer (1=yes) |
| 7 | | Tloc spill guts? answer (no) (Provides Diagnostics) |
| 8 | | Fragmentography run? answer (no) |
| 9 | nnnnnn | Minimum for less than 5 peaks (6 digits) |
| 10 | nnnnnn | Minimum for more than 4 peaks (6 digits) |
| 11 | | New value for points left of node (default value) |
| 12 | | New value for points right of node (default value) |
| 13 | | New minimum thirds (default value) |
| 14 | | New detection window (default value) |
| 15 | | New rate minimum (default value) |
| 16 | .nnnn | New linear ratio (value from BATCHJOB question) |
| 17 | | IS THIS FOR AN ISOMER? answer (no) |
| 18 | | SHORTEN SCAN INTERVAL? answer (no) |
| 19 | | MONITOR THE Peak RESOLVERS? answer (no) |

< The following records are used by other GC/MS programs.>

Record 20:
AERL Tape number and file number (2I4)

Record 21:
Fraction type (6A1)
3 sets: Internal standard (20A1)
Concentration (3A1)
Concentration factor (4A1)
Injection volume (A1)

Record 22:
Analytical Laboratory (25A1)
GC Column data (28A1)

Record 23:
Industrial category (2A1)

** NOTE: The first 19 records are the input for CLEANUP.

**INPUTS**

1. TAPE (FROM TAPE CONVERSION PROGRAMS)
2. FOR005.DAT OR INTERACTIVE TERMINAL
   (CONTROLLING FILE)
3. UNITS.DAT
   (LUN'S FOR TERMINAL INPUT & OUTPUT)

---

**PROGRAM MODULE**

RETRIEVE

---

**OUTPUTS**

1. <RUNNAME>.FIL
   (RAW DATA FILE OF MASS/INT. PAIRS)
2. <RUNNAME>.TIM
   (FILE CONTAINING RUN PARAMETERS)

**INPUTS**

1. <RUNNAME>.FIL
   (RAW DATA FILE OF MASS/INT. PAIRS)
2. FOR005.DAT OR INTERACTIVE TERMINAL
   (CONTROLLING FILE)
3. UNITS.DAT
   (LUN'S FOR TERMINAL INPUT & OUTPUT)

**PROGRAM MODULE**

BUILD

**OUTPUTS**

1. <RUNNAME>.DAT
   (DIRECT ACCESS FILE OF MASS/INT. PAIRS)
2. <RUNNAME>.PAR
   (DIRECT ACCESS FILE CONTAINING PARAMETER INFORMATION ABOUT THE <RUNNAME>.DAT FILE)

**INPUTS**

1. <RUNNAME>.TIM
   (FILE CONTAINING RUN PARAMETERS)
2. <RUNNAME>.PAR
   (DIRECT ACCESS FILE CONTAINING PARAMETER INFORMATION ABOUT THE <RUNNAME>.DAT FILE)
3. <RUNNAME>.DAT
   (DIRECT ACCESS FILE OF MASS/INT. PAIRS)
4. FOR005.DAT OR INTERACTIVE TERMINAL
   (CONTROLLING FILE)
5. UNITS.DAT
   (LUN'S FOR TERMINAL INPUT & OUTPUT)
6. MASS.STD
   (SELECTED MASSES FOR USE IN BACKGROUND SUBTRACTION)

**PROGRAM MODULE**

STANDARD

**OUTPUTS**

1. <RUNNAME>.BCK
   (DIRECT ACCESS FILE OF BACKGROUND SUBTRACTED MASS/INT. PAIRS)
2. MASS.TMP
   (TEMPORARY FILE OF MASSES FOR INTERNAL STANDARD LOCATION)

**INPUTS**

1. <RUNNAME>.PAR
   (DIRECT ACCESS FILE CONTAINING PARAMETER INFORMATION ABOUT THE <RUNNAME>.DAT FILE)
2. <RUNNAME>.BCK
   (DIRECT ACCESS FILE OF BACKGROUND SUBTRACTED MASS/INT. PAIRS)
3. MASS.TMP
   (TEMPORARY FILE OF MASSES FOR INTERNAL STANDARD LOCATION)
4. FOR005.DAT OR INTERACTIVE TERMINAL
   (CONTROLLING FILE)
5. UNITS.DAT
   (LUN'S FOR TERMINAL INPUT & OUTPUT)

**PROGRAM MODULE**

PEAK

**OUTPUTS**

1. <RUNNAME>.STD
   (FILE CONTAINING INTERNAL STANDARD(S) LOCATION AND AREA UNDER BASE PEAK)

**INPUTS**

1. <RUNNAME>.DAT
   (DIRECT ACCESS FILE OF MASS/INT. PAIRS)
2. <RUNNAME>.PAR
   (DIRECT ACCESS FILE CONTAINING PARAMETER INFORMATION ABOUT THE <RUNNAME>.DAT FILE)
3. FOR005.DAT OR INTERACTIVE TERMINAL
   (CONTROLLING FILE)
4. UNITS.DAT
   (LUN'S FOR TERMINAL INPUT & OUTPUT)

---

**PROGRAM MODULE**

MODELS

---

**OUTPUTS**

1. <RUNNAME>.MOD
   (MODEL PEAK INFORMATION FOR CLEANUP)

**INPUTS**

1. <RUNNAME>.PAR
   (DIRECT ACCESS FILE CONTAINING PARAMETER INFORMATION ABOUT THE <RUNNAME>.DAT FILE)
2. <RUNNAME>.DAT
   (DIRECT ACCESS FILE OF MASS/INT. PAIRS)
3. <RUNNAME>.MOD
   (MODEL PEAK INFORMATION FOR CLEANUP)
4. UNITS.DAT
   (LUN'S FOR TERMINAL INPUT & OUTPUT)
5. FOR005.DAT OR INTERACTIVE TERMINAL
   (CONTROLLING FILE)
6. LIST.MAS
   (TEMPORARY WORK FILE)

**PROGRAM MODULE**

CLEANUP

**OUTPUTS**

1. <RUNNAME>.CLN
   (PEAK LOCATION, AREA UNDER PEAK, AND MASS/INT. PAIRS FOR PEAK)

**INPUTS**

1. UVALS.DAT
   (U VALUES FOR MASSES 1 – 400)
2. <RUNNAME>.CLN
   (PEAK LOCATION, AREA UNDER PEAK, AND MASS/INT. PAIRS FOR PEAK)

**PROGRAM MODULE**

PBMUNK

**OUTPUTS**

1. INPUT.PBM
   (DIRECT ACCESS FILE OF CONDENSED UNKOWN SPECTRA FOR INPUT TO PBM)

**INPUTS**

1. INPUT.PBM
   (DIRECT ACCESS FILE OF CONDENSED UNKNOWN SPECTRA FOR INPUT TO PBM)
2. FOR005.DAT OR INTERACTIVE TERMINAL INPUT
   (CONTROLLING FILE)
3. UNITS.DAT
   (LUN'S FOR TERMINAL INPUT & OUTPUT)

---

**PROGRAM MODULE**

PBM

---

**OUTPUTS**

1. <RUNNAME>.PBM
   (DIRECT ACCESS FILE OF PBM MATCHES FOR EACH PEAK)

**INPUTS**

1. <RUNNAME>.STD
   (FILE CONTAINING INTERNAL STANDARD(S)
   LOCATION AND AREA UNDER BASE PEAK)
2. <RUNNAME>.TIM
   (FILE CONTAINING RUN PARAMETERS)
3. <RUNNAME>.PBM
   (DIRECT ACCESS FILE OF PBM MATCHES
   FOR EACH PEAK)
4. <RUNNAME>.CLN
   (PEAK LOCATION, AREA UNDER PEAK,
   AND MASS/INT. PAIRS FOR PEAK)
5. INTSTD.DAT
   (LIST OF VALID INTERNAL STANDARDS)
6. GCCOL.DAT
   (LIST OF VALID GC COLUMNS)
7. ANALAB.DAT
   (LIST OF VALID ANALYTICAL LABORATORIES)
8. FOR005.DAT OR INTERACTIVE TERMINAL
   (CONTROLLING FILE)
9. UNITS.DAT
   (LUN'S FOR TERMINAL INPUT & OUTPUT)

- - - - - - - - - - - - - - - - - - - - - - - - -

**PROGRAM MODULE** TIMSEK

- - - - - - - - - - - - - - - - - - - - - - - - -

**OUTPUTS**

1. <RUNNAME>.SEK
   (DIRECT ACCESS FILE CONTAINING RUN
   INFORMATION AND PBM MATCHES, MASS/INT.
   PAIRS, RRT, AND CONCENTRATION FOR
   EACH PEAK FOUND BY CLEANUP)
2. AREALTIC.STD
   (A FILE CONTAINING A RECORD OF WHERE
   AND WHO FOUND THE INTERNAL STANDARD)

**INPUTS**

1. <RUNNAME>.SEK
   (DIRECT ACCESS FILE CONTAINING RUN INFORMATION AND PBM MATCHES, MASS/INT. PAIRS, RRT, CONCENTRATION FOR EACH PEAK FOUND BY CLEANUP)
2. GCCOL.DAT
   (LIST OF VALID GC COLUMNS)
3. NAMES.DIR
   (DIRECT ACCESS FILE CONTAINING COMPOUND NAMES, CAS NO., MOLECULAR WEIGHT, AND MOLECULAR FORMULA FOR EACH RECORD IN PBM DATABASE)
4. UNITS.DAT
   (LUN'S FOR TERMINAL INPUT & OUTPUT)

- - - - - - - - - - - - - - - - - - - - - - - - - - -

**PROGRAM MODULE**

HISLIB

- - - - - - - - - - - - - - - - - - - - - - - - - - -

**OUTPUTS**

1. <RUNNAME>.HLD
   (TEMPORARY HOLDING FILE OF PEAKS FOR CHEMISTS(S) VERIFICATION. SAME FORMAT AS THE <RUNNAME>.SEK)
2. <RUNNAME>.MIS
   (TEMPORARY HOLDING FILE OF PEAKS TO BE PROCESSED BY THE MISS ALGORITHM. SAME FORMAT AS THE <RUNNAME>.SEK)
3. HISLIB.HIT & RRT.HIT
   (HISTORICAL LIBRARY)
4. TRANSACT.HIT
   (TRANSACTION FILE FOR LIBRARY OF TENTATIVELY IDENTIFIED COMPOUNDS)
5. TRANSACT.MIS
   (TRANSACTION FILE FOR LIBRARY OF UNIDENTIFIED SPECTRA)

**INPUTS**

1. <RUNNAME>.CLN
   (PEAK LOCATION, AREA UNDER PEAK, AND MASS/INT. PAIRS FOR PEAK)
2. FOR005.DAT OR TERMINAL INPUT
   (CONTROLLING FILE)
3. UNITS.DAT
   (LUN'S FOR TERMINAL INPUT & OUTPUT)

- - - - - - - - - - - - - - - - - - - - - - - - -

**PROGRAM MODULE**

SPECTRA

- - - - - - - - - - - - - - - - - - - - - - - - -

**OUTPUTS**

1. <RUNNAME>.PLX
   (PLOT FILE OF EACH PEAK)

**INPUTS**

1. <RUNNAME>.FIL
   (RAW DATA FILE OF MASS/INT. PAIRS)
2. UNITS.DAT
   (LUN'S FOR TERMINAL INPUT & OUTPUT)
3. FOR005.DAT OR INTERACTIVE TERMINAL
   (CONTROLLING FILE)

- - - - - - - - - - - - - - - - - - - - - - - - -

**PROGRAM MODULE**

GCPLOT

- - - - - - - - - - - - - - - - - - - - - - - - -

**OUTPUTS**

1. <RUNNAME>.PLX
   (PLOT FILE OF RECONSTRUCTED
   GAS CHROMATOGRAPH)

# ADDITION OF SPECTRUM TO THE PBM DATABASE

The module PBMADD adds spectra to the end of the PBM database. This module is essentially the same program as PBM condensation program (PBMCON), except that it adds spectra to the end of a current database instead of creating a new one. The following files are required:

DB0:NAMES.DIR - File of corresponding compound names
DB0:PBMFIL.DAT - PBM spectral database
DB0:PBMEND.DAT - Contains that record number of the last spectrum in the database
DB0:UVALS.DAT - U + A values for masses 1-400
DB0:SPEC.DAT - Input file of spectra to be added to database.

The module is executed by issuing the following command:

```
PDS> RUN PBMADD
```

The following command will build the task image:

```
PDS> @PBMADD
```

## FILE STRUCTURE

SPEC.DAT

INPUT FILE OF SPECTRA THAT
ARE TO BE ADDED TO THE PBM
DATABASE.
SEQUENTIAL FILE

## HEADER RECORDS FOR SPECTRUM

### RECORD 1

```
   Compound name - 72 characters
   READ(N,10)(CMPNAM(I),I=1,72)
10 FORMAT(72A1)
```

### RECORD 2

```
   Molecular formula - 20 characters
   READ(N,15)(FORMUL(I),I=1,20)
15 FORMAT(20A1)
```

### RECORD 3

```
   Molecular weight & Cas #
   READ(N,20)WEIGHT,ICAS
20 FORMAT(I3,1x,I10)
```

### RECORD 4

```
   Molecular composition array
   READ(N,25)(MOCMP(I),I=1,12)
25 FORMAT(12I2)
```

Each element of the array (except the last one) corresponds to a count of the 11 most common chemical elements if they are part of the molecular formula and zero if absent.

| Chemical element | C | H | BR | CL | F | I | N | O | P | S | SI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Array element | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |

The twelfth element of the array corresponds to a count of chemical elements that are not among the 11 most common.

Example:

D10 ANTHRACENE
C14 D10

| Array element | C | H | BR | CL | F | I | N | O | P | S | SI | OTHER |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Count of chemical element | 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |

## RECORDS FOR MASS/INT PAIRS OF SPECTRUM

N NUMBER OF RECORDS
13 MASS/INT PAIRS/RECORD

```
   READ(N,30)(MAS(I),INTEN(I),I=1,13)
30 FORMAT(13(2I3))
```

NOTE - A mass intensity of zero signifies the end of the mass/inten pairs. The next records contains the first header record of the next spectrum.

NOTE - The mass value is an integer in the range of 1 to 999, and the intensity value represents a normalized integer intensity value in the range from .01% to 99.9%. (the decimal point is omitted from the value when entering the spectrum.)

# APPENDIX II.

# APPLICATIONS OF A DATA ADAPTIVE BACKGROUND SUBTRACTION TECHNIQUE TO THE GC/MS ANALYSIS OF WASTE WATER SAMPLES

by

A. D. Sauter[1], L. Faas, and G. R. Kurth
Computer Sciences Corporation
Falls Church, Virginia

and

W. M. Shackelford and D. M. Cline
U.S. Environmental Protection Agency
Environmental Research Laboratory
Athens, Georgia

---

[1] Present address - U.S. Environmental Protection Agency, Environmental Monitoring and Support Laboratory, Las Vegas, NV

# ABSTRACT

In the GC/MS analysis of samples containing many organic compounds, interferences from the sample matrix along with instrumental noise can severely limit the ability to accurately determine the intensity of an ion of given m/z value. Lack of accuracy in this determination can result in erroneous spectrum identification and component quantification. Because much of the routine GC/MS analysis being performed upon environmental samples relies upon internal standard methods for quantification, accuracy in determining the absolute intensities of the ions used for quantification is essential. In this work, a data adaptive background subtraction algorithm was applied to complex extracted ion current profiles (EICP) corresponding to key ions in the GC/MS. Computer programs were written to calculate a background-free EICP for as many as 30 ions concurrently. The resultant EICP sets are then used in place of the raw data for compound specific or class specific reverse searches. The peak areas calculated after background processing are shown to be within the precision limits of those calculated after applying manual techniques of background compensation. An increase in the ability of automated feature extraction techniques to recognize ions subject to severe interferences is demonstrated.

# INTRODUCTION

In recent years gas chromatography/mass spectrometry (GC/MS) has become a routine tool for the analysis of wastewater for organics. The U.S. Environmental Protection Agency (EPA) in its recent study of industrial waste treatment technology relied heavily on GC/MS techniques to find and quantify organic priority pollutants in industrial wastewater samples.[1-3] In this study, compounds of interest were identified through both mass spectrometric (MS) data and retention time data. Internal standards were used in these analyses to provide reference data both for quantification and identification. Use of retention indices coupled with MS data for identification purposes was reported by Nau and Biemann[4] and Sweeley et al[5]. Subsequently, automated systems in several laboratories[6-9] have been reported that rely upon retention data as well as MS data for computer aided identification of a known set of organic compounds.

As noted by Nau and Biemann[4] the use of internal standards for qualitative and quantitative purposes within an analysis necessitates that the internal standard be able to be unambiguously located and its intensity be able to be measured accurately. In analyzing highly predictable matrices, one can choose internal standards that have few interferences from the background and are present in sufficient quantity to be adequately measured. In cases where the internal standard has matrix interference, problems arise quickly.

This work resulted from problems encountered in a computer survey[9] of thousands of GC/MS analyses acquired in 18 different laboratories on a variety of GC/MS instrumentation under similar conditions. Common among all analyses were internal standards that could be used to compare the retention data of the various compounds and to provide a means for quantification. Here, as in the systems cited above, finding the internal standard was an indispensible step in the analysis. Both quantification and identification depended on the ability of an operator or a computer to accurately gather MS data and retention data in the presence of varying degrees of interference.

In the initial phases of the computer survey, it became obvious that locating the internal standards was beyond the capability of the automated system in some cases. For example, the sample analysis procedure called for adding the internal standard at a concentration of 20 μg/ml regardless of sample complexity. In cases where 20 μg/ml was an extremely minor concentration, the computer system lacked the sensitivity to extract a spectrum that the spectrum matching program could identify. Consequently, work in this case had to cease until the internal standard could be located manually. Also, the superpositioning of ion signals from the internal standard on either a noisy baseline or a sharply rising or falling baseline made spectrum extraction and subsequent matching haphazard and made the computer survey in this case impossible without manual assistance.

To remedy this situation, a preprocessor to the survey system was constructed to look specifically for the internal standard. The ion current profiles of key internal standard ions were examined for simultaneous peaking and at such a point, provided the relative abundances met certain criteria, the internal standard was said to be located. Although this remedy eliminated many problems associated with sensitivity, baseline problems continued to hamper peak recognition and relative abundance calculation.

An automated data-adaptive solution to some of these problems has been published by Summons et al.[10] and was used to form the basis of our approach. In the earlier work, data were collected under carefully controlled conditions in one laboratory. Because our data were received from many sources and acquired from a variety of sample matrices, the work by Summons et al.[10] was modified to make the system even more flexible and data-adaptive.

These modifications, as well as the performance of the automated system on field data are presented for several cases of data complexity. The applicability of this technique to reverse search analysis is also demonstrated.

## EXPERIMENTAL

Instrumentation. All data were acquired on commercially available integrated GC/MS/data systems. As shown in Table I, 18 different systems in 14 laboratories were involved in the study.

TABLE I. INSTRUMENTATION FOR GC/MS SURVEY WITH CONTRIBUTING LABORATORIES

1. Finnigan 3200 MS, Varian 1200 GC, System Industries data system--Region IV (EPA), Region II (EPA), Region VII (EPA)
2. Finnigan 4000 MS, Finnigan 9600 GC, Finnigan-Incos data system--Region II (EPA), Region IV (EPA), Region VII (EPA), Carborundum Co., West Coast Technical Service, Jacobs Engineering, Analytical Research Laboratories, Inc.
3. Dupont Dimaspec MS, Finnigan 9600 GC, Finnigan-Incos data system--Systems, Science, and Software; TRW
4. Hewlett-Packard 5985 GC/MS/data system--Radian Corp., Gulf South Research Institute, Envirodyne Engineers
5. Hewlett-Packard 5982 GC/MS/data system--Radian Corp., Foremost Foods, Energy Research Corp.

Computer programs were run on a Digital Electronics Corporation (DEC) PDP 11/70 computer with 256K words of memory, 2K words of cache memory, two 88 M-byte disk drives (DEC RP04), two 9-track magnetic tape drives (DEC TU16), a Calcomp 936 plotter and a Tektronics 4014 display terminal. Statistical analyses were performed using the Biomedical Computer Program P-Series package (University of California Press, 1979.)

Laboratory Procedures. Data submitted to this laboratory for study consisted of raw GC/MS data acquired in a survey of industrial waste water. The procedures used for analysis were recently published[3]. A purgeable fraction and base/neutral and acid extractable fractions were analyzed by GC/MS for each sample taken in the field. One internal standard, anthracene-$d_{10}$, was added to the final extract of the acid and base/neutral fraction. Two internal standards, bromochloromethane and 1,4-dichlorobutane, were added to the purgeable fraction. During the whole of the study, one GC column liquid phase was used for purgeable analysis (0.2% Carbowax 1500), two liquid phases were used for acids analysis (Tenax GC and 1% SP-1240 DA); and three liquid phases were used for base/neutrals analysis (1% SP-2250, 3% SP-2250, 3% SP-2250 DB). Column performance was assured by the chromatography of test compounds (benzidine, pentachlorophenol) and GC/MS performance was required to meet specifications set forth by Eichelberger et al.[11]

Data Analysis Procedure. A complete description of the data analysis system and procedures for analysis can be found elsewhere[9]. The procedure used for specific ion searches, whether for internal standard location, target compound location, or compound class searching, involves both manual and automated operations. The amount of manual intervention called for depends upon the degree to which the algorithms are tailored to each particular sample run. In routine cases, parameters were chosen that had been shown to give generally good results for the widest variety of run types. For special cases, such as a class specific search for polychlorinated biphenyl moieties, program parameters are varied until the best fit of the data is obtained.

Background subtraction. The background correction technique as reported by Summons et al.[10] utilizes a "piecewise" least squares fit approach to estimate the local background in spectrum number ranges. In our work, extracted ion current profiles (EICP) are divided into sequences of N intensity values. Each sequence is further divided into intervals. The number of intervals, NUMINT, and the number of spectra in an interval, NUMSCAN, are user defined variables employed in this technique. The product of these variables is the sequence size.

A program operates on raw data as follows. For each EICP sequence, the lowest intensity in each interval is determined. A least squares line is calculated for these points within each sequence. Next all intensity values that fall above this first least squares lines are eliminated, and a new line is calculated with the remaining data points. In this way points that are not on the base line such as those between unresolved multiplets are eliminated from background consideration.

Summons et al.[10] employed a spline function to estimate the background between adjacent sequence midpoints (equation 1).

$$f(\varepsilon) = 1-3\varepsilon^2+2\varepsilon^3 \tag{1}$$

This joining function is evaluated at each spectrum between the midpoints of adjacent sequences by equation 2 where $\varepsilon$ is the

$$\varepsilon = \frac{t - t_{mid}\ (i)}{t_{mid}\ (i+1) - t_{mid}\ (i)} \tag{2}$$

scan number of interest in this range and $t_{mid}$ (i) and $t_{mid}$ (i+1) are the midpoint scan or spectrum numbers of the respective sequences $S_I$ and $S_{I+1}$. Figure 1 is a graphic repre-

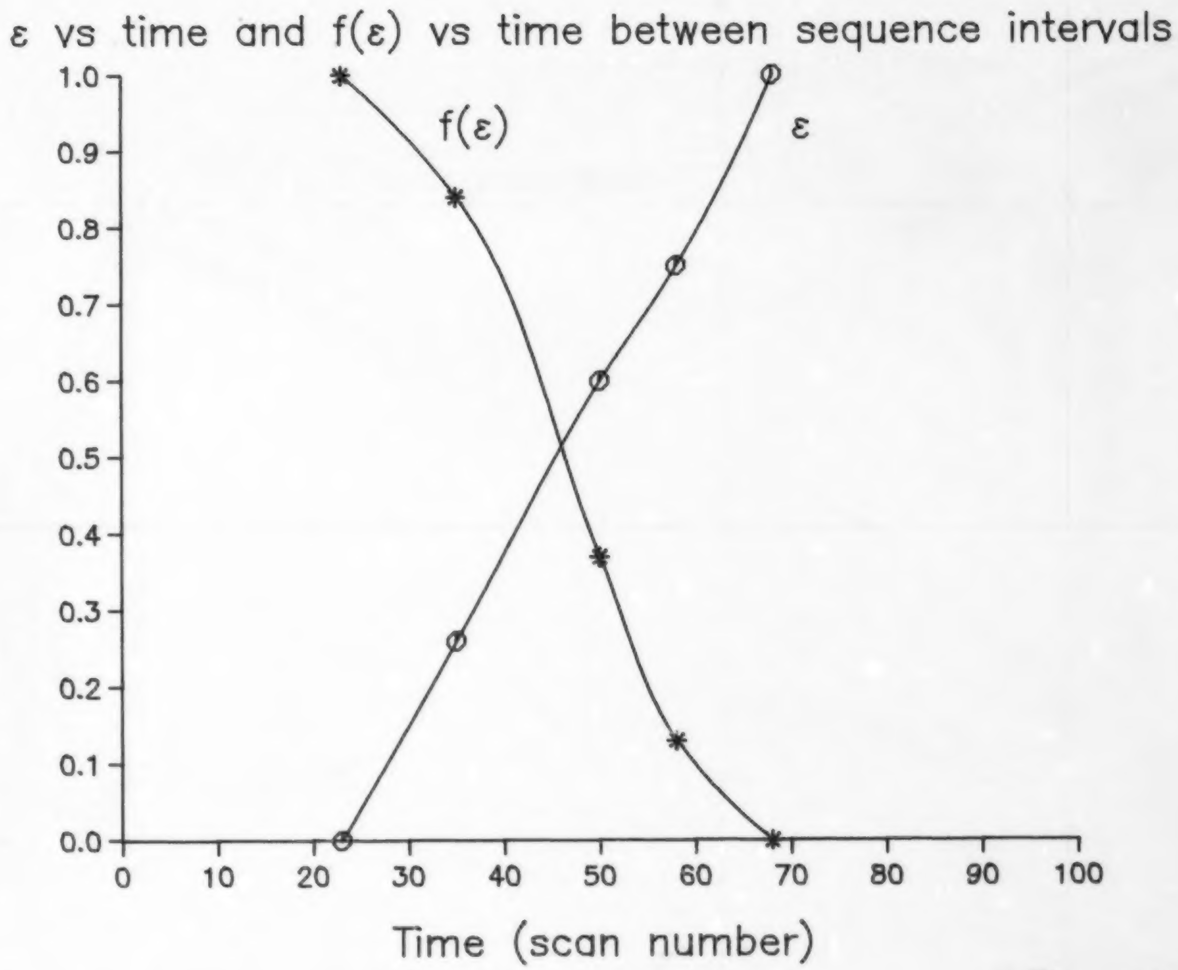


Figure 1. Plot of f(ε) vs. time for equation 1 and ε vs. time for equation 2.

sentation of equations 1 and 2. The background intensity is then calculated by equation 3.

$$B(t) = f(\varepsilon)\ P_i\ (t) + [1-f(\varepsilon)]\ P_{i+1}\ (t) \qquad (3)$$

where $P_i(t)$ and $P_{(i+1)}$ (t) are the second least square approximations for sequence I and I-1.

The background intensities [B(t)] are then subtracted from the raw data. Simple and complex EICP studied initially gave results that were encouraging because the ion current removed in most cases agreed well with manually estimated background values.

An example of a simple case is shown in Figure 2 with both raw and background-corrected EICP presented. The peak height was manually estimated for the raw data at approximately 1990 intensity units. The peak height resulting after automated background subtraction (and before thresholding) was 2030 intensity units. The agreement between manual and automatic background subtraction techniques for this simple EICP is representative of our observations. This example serves to demonstrate that the technique produces results that are consistent with manual compensation. The power of this approach becomes apparent when complex EICP are studied.

Figure 3 shows a complex EICP. By inspection, it is apparent that the relative intensities of peak maxima were preserved along with EICP fine structure. To demonstrate that manual and automatic background-corrected intensities are in agreement, manual peak height measurements are compared to background corrected data in Table 2 for the eight peaks labeled in Figure 3. The manual data represents the average of four chemists' estimation of peak heights above background. The automated background compensation routine generally falls well within one standard deviation of the corresponding manual value. The apparent positive bias of the automated values is probably due to the fact that the background line is constructed through peak minima.

Table 2 indicates that the automated background subtraction technique can produce results consistent with manual background corrected estimation of peak heights. The EICP in Figure 3 contains resolved and unresolved peaks of varying peak shapes and consequently represents a difficult test. Thus, it can be seen that the automated background subtraction technique can accurately compensate for the fraction of the ion current due to "background" for major peaks in a given complex EICP. Minor peaks in clusters have not undergone background correction in many cases, and as such, represent a limitation of the background correction technique as applied to data that we encoun-

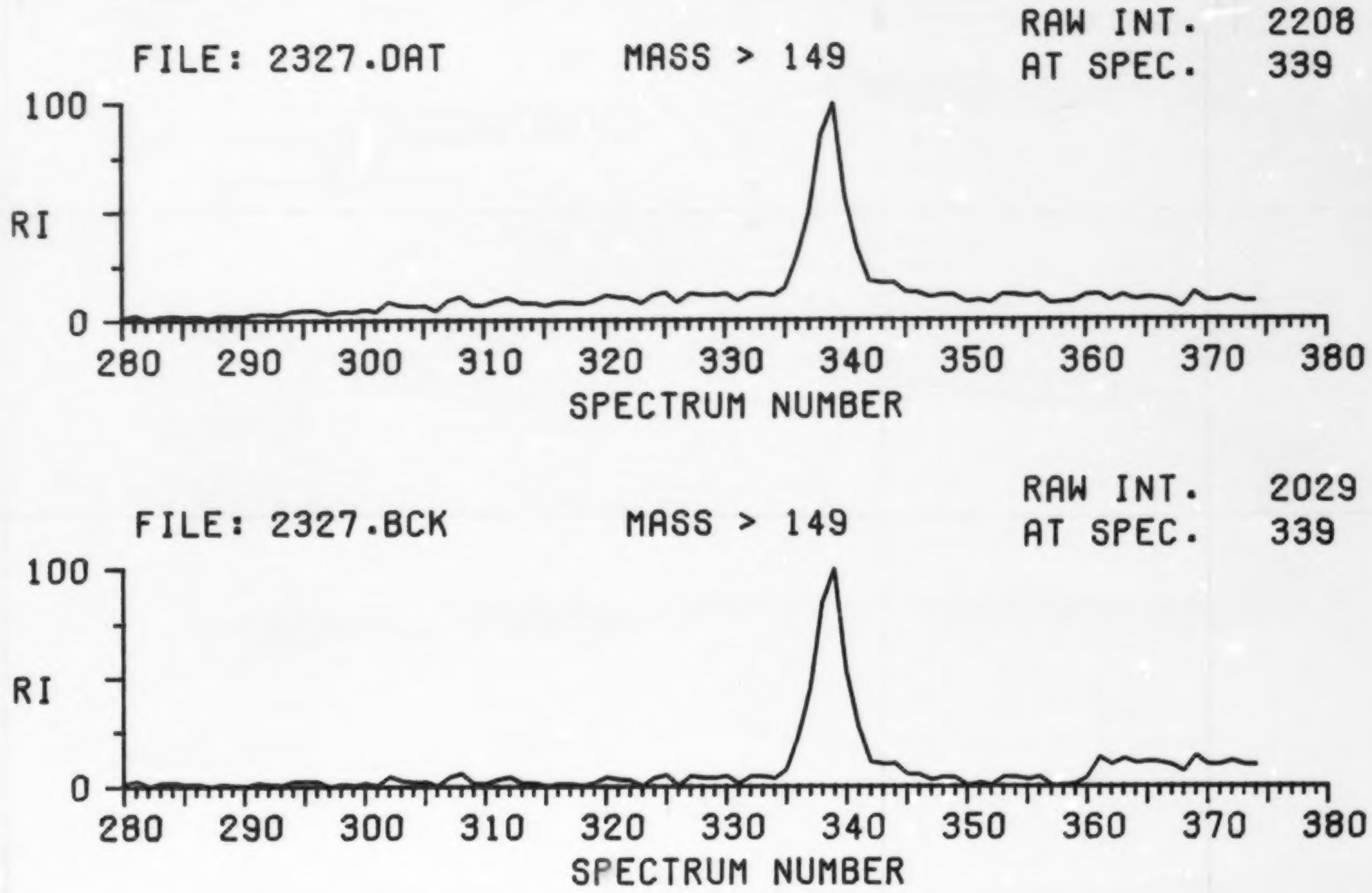


Figure 2. Background subtraction for the simple case of an isolated large peak. The upper plot depicts raw data.

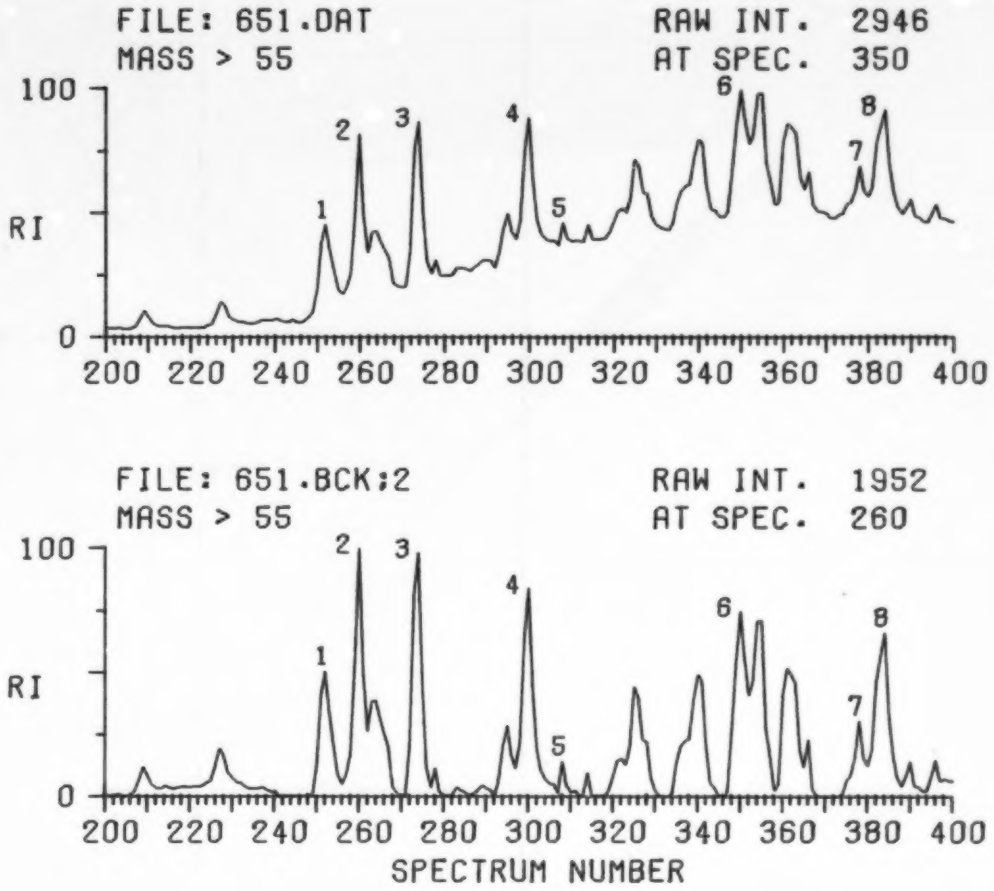


Figure 3. Background subtraction for a complex case.
The upper plot depicts raw data.

tered in our work. Oftentimes in practice this does not represent a major limitation because identification of species of relatively low concentration (in the work studied[9] less than 10 ppb) are of limited interest. Adjusting the user-defined variables can compensate to a degree for this property of the background correction program.

TABLE 2. CALCULATED PEAK HEIGHTS

| Peak Number* | Peak Height<br>Manual** | <br>Automatic | Δ |
|---|---|---|---|
| 1 | 1050 ± 36 | 1045 | - 5 |
| 2 | 1932 ± 49 | 1947 | +15 |
| 3 | 1902 ± 56 | 1909 | + 7 |
| 4 | 1621 ± 35 | 1636 | +15 |
| 5 | 261 ± 31 | 273 | +12 |
| 6 | 1462 ± 62 | 1462 | 0 |
| 7 | 597 ± 19 | 591 | - 6 |
| 8 | 1253 ± 15 | 1268 | +15 |

* Refer to Figure 1.

** Average of observations by 4 chemists.

The optimum value of the user defined variables (NUMSCAN and NUMINT) is peak geometry dependent. For example, when NUMSCAN is lowered to 3 (spectra/interval) and NUMINT is raised to 15 (intervals/sequence) maintaining a constant sequence size (45 spectra), the broad peak (# 1) in Figure 4 is too aggressively subtracted. The area for this peak is only 90% of the value attained when NUMSCAN and NUMINT are set to five and nine, respectively, and only 92% of the chemist estimated value. Some sharp peaks (shown in Figure 3), however, showed an equivalent or improved agreement with the manually estimated peak height with these values of the user defined parameters.

Another example of the effect of the user defined variables NUMINT and NUMSCAN is illustrated below. Figure 5 shows the background correction technique applied with default values for NUMSCAN and NUMINT (5 and 9). The manually estimated peak height (5A) for the raw data was 570 intensity units. Figure 5B has a chopped appearance and the peak height (474) is considerably lower than the manually estimated value. By changing NUMSCAN and NUMINT to values of 10 and 8 the resultant peak height was increased to 575 intensity units (5C), which is in better agreement with the manual peak height estimation. This observation served to illustrate that peak width considerations can affect the results of this technique even for simple EICP. Because EICP can contain peaks or clusters of varied geometry and complexity, it is unlikely that any one fixed

BEGIN FIRST LINE OF TEXT HERE
TYPING GUIDE SHEET
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD, BEGIN SECTION HERE

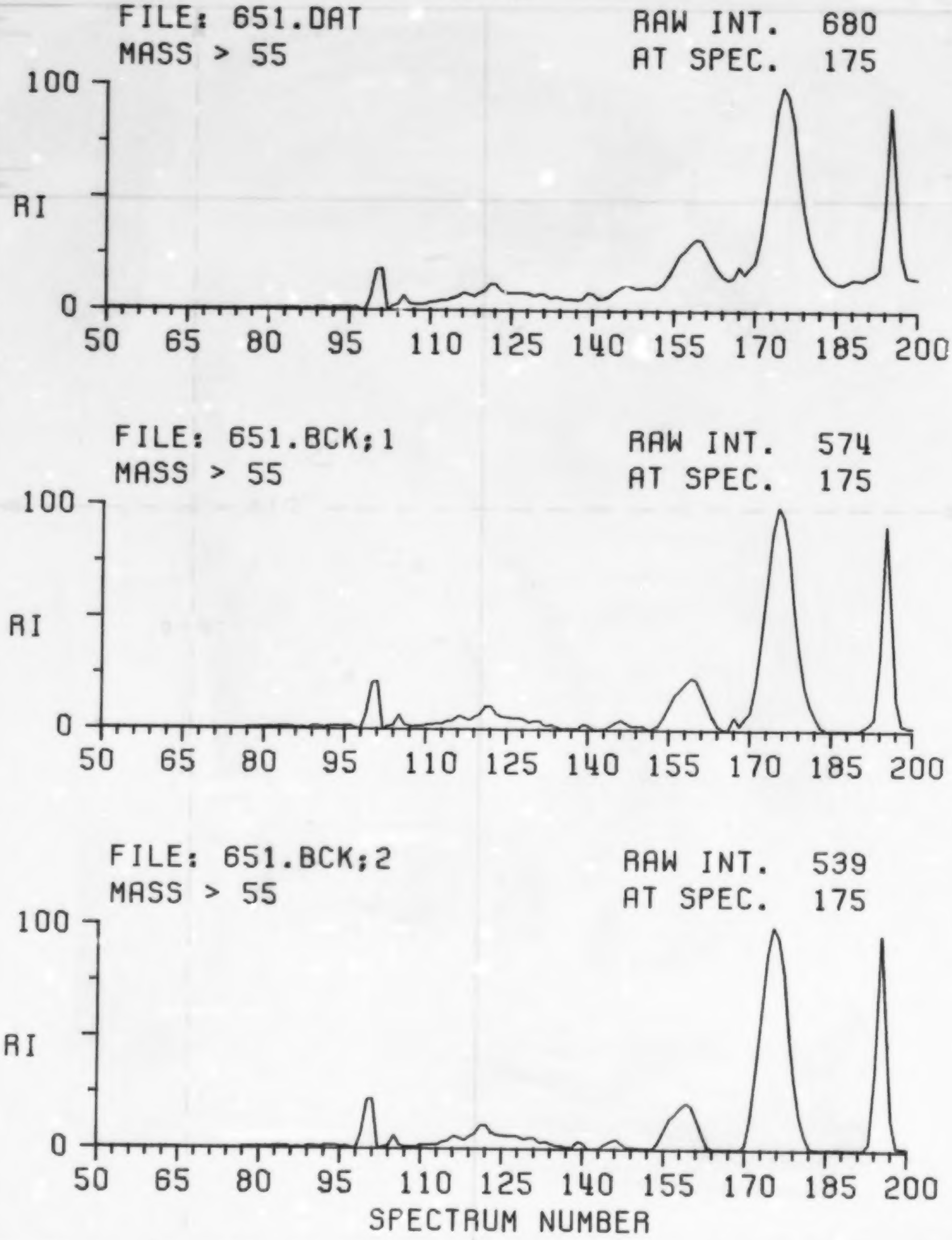


Figure 4. Effect of NUMSCAN/NUMINT on background subtract. The top plot is uncorrected data; the middle is data corrected with NUMSCAN/NUMINT set at 5/9; the bottom plot slows data corrected with NUMSCAN/NUMINT set to 5/15.

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS

PAGE NUMBER
EPA-287 (Cin.)
(4-76)

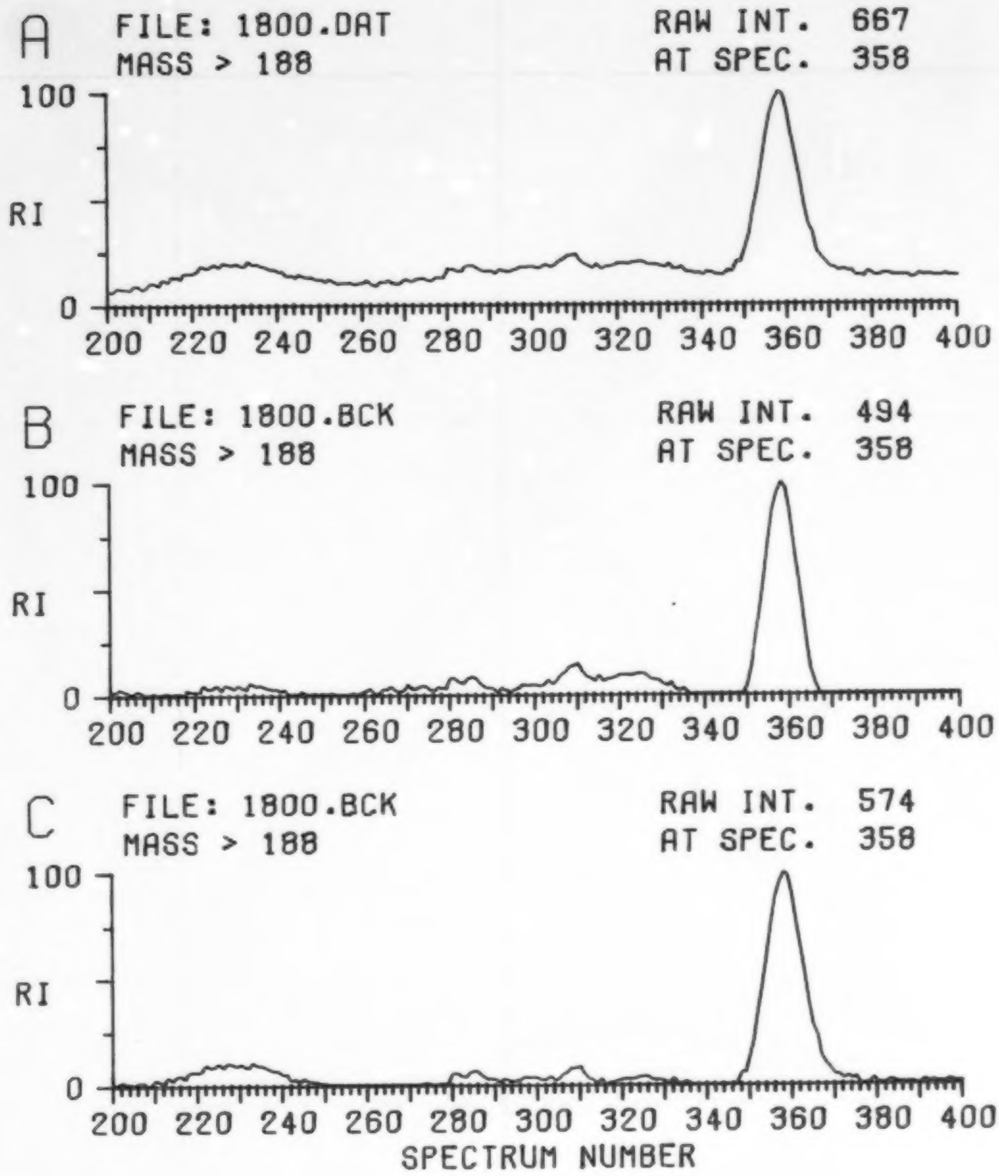


Figure 5. Effect of NUMSCAN/NUMINT on background subtract. The top plot is uncorrected; the middle plot is corrected with NUMSCAN/NUMINT set at 5/9; the bottom plot is corrected with NUMSCAN/NUMINT set at 10/8.

value for the user-defined variables NUMSCAN/NUMINT will be adequate for all peaks in a complex EICP. Because our initial application of the background subtraction technique was in the location of internal standards, NUMSCAN/NUMINT values were chosen that gave background corrected peak heights in agreement with manually estimated peak heights for major fragments in the mass spectrum of the internal standards in simple EICP. For data collected on GC columns except Tenax GC, the NUMSCAN/NUMINT values were set to 5/9. For Tenax GC values of NUMSCAN/NUMINT were set at 10/8.

Noise thresholding. After the best background line has been removed from the raw data, detection of the peak representing the component is initiated. In the work of Summons et al.[10] an intensity threshold was set to compensate for noise peaks that were above the corrected background line. The intensity threshold was set by examining the distribution of intensities within the background corrected data set. A data adaptive threshold was calculated by

$$\text{Threshold} = \text{Mode} + 2.5 \times (\text{standard deviation from the mode}) \qquad (4)$$

and appeared to give excellent noise compensation. In the present application, however, this approach to thresholding gave high values for the threshold that resulted in some peaks of interest falling below the threshold value. Our data did not exhibit good noise statistics because the usual mode of acquiring data involves setting an instrumental threshold for the noise. Thus, the noise looks like spikes on a flat baseline rather than a sinusoidal variance about some mode value. Although the peaks of the noise spikes may vary in height, the minima are invariably equal to zero. As the uncorrected baseline rises above the abcissa, use of the least squares technique puts some minima above the calculated baseline and some below it. The variance of these points about the calculated baseline not only involves instrument noise but also useable data. Thus, in our work the standard deviation from the mode value was too great to use for a signal threshold.

Intensity distributions of background-corrected EICP were examined to derive an alternate threshold setting procedure. A typical ranged intensity distribution for one of the columns tested is shown in Figure 6 (raw and background corrected EICP are shown also in Figure 6). This ranged intensity distribution is representative of the positively skewed distributions that resulted after background subtraction of simple EICP. The degree to which the raw data influenced these background corrected intensity distributions was not delineated; but it was recognized that the raw data, the algorithm employed, and other factors could predispose the skew of the distribution

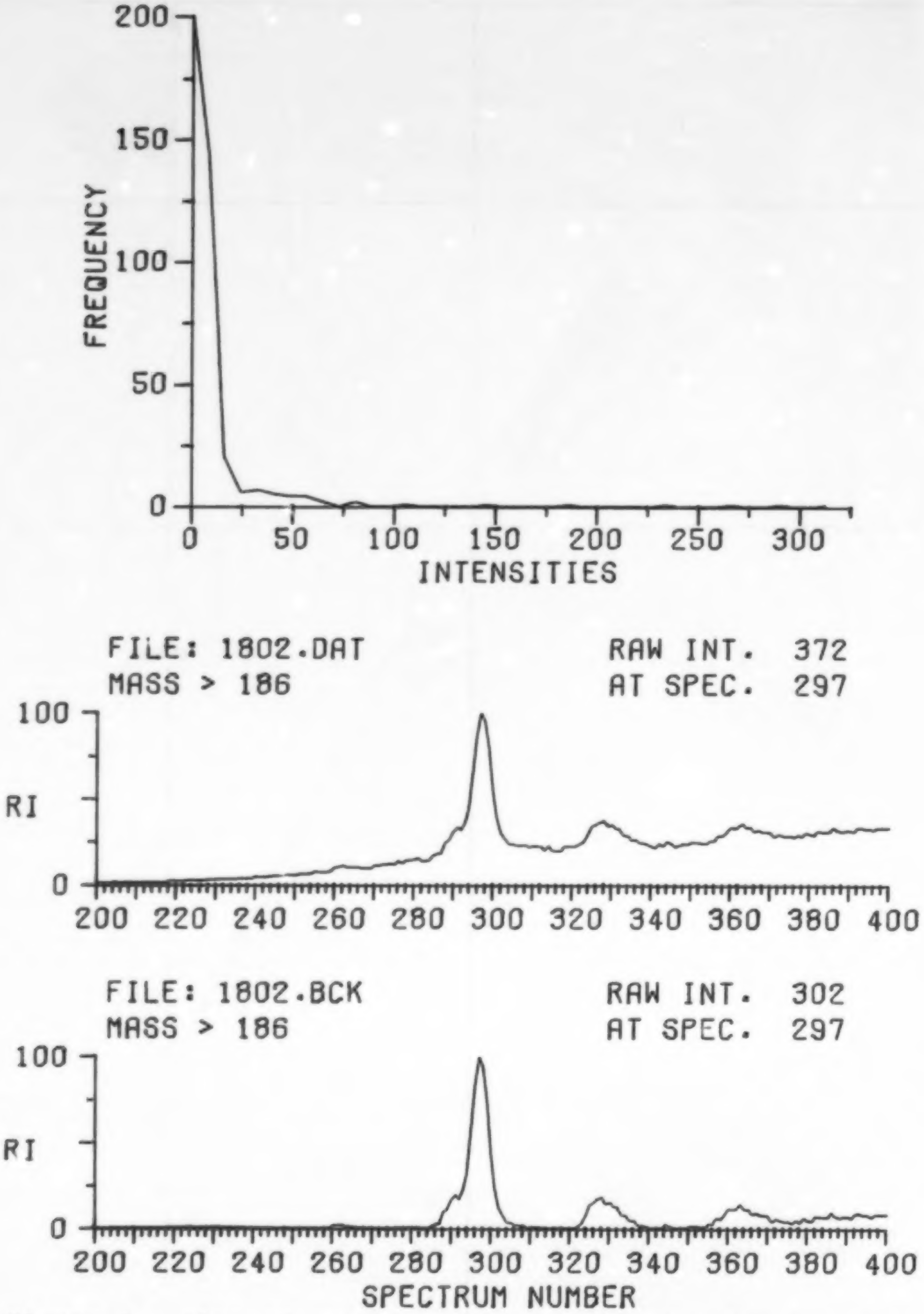


Figure 6. Plot of ranged intensity distribution for mass 186 after background correction. Chromatograms show uncorrected (upper) and background corrected (lower) data.

resulting from background correction. For example, when many zero intensity values occur in the raw data of any EICP, the background corrected EICP intensity distribution is likely to have a mode of zero intensity units. The degree to which these distributions represent the effect of the background correction technique as opposed to the raw data characteristics prior to background substraction is a question that we did not address quantitatively. Empirically, we found that simple EICP yielded a positively skewed distribution with a zero value mode.

Other distribution types were noted also. For example, when ion current fluctuations around a given discrete value (above zero) were apparent in the background-corrected EICP, the distribution had a higher degree of symmetry and was generally less peaked and gave modes at greater than zero intensity units. A partial EICP and the total intensity distribution resulting after background correction is shown in Figure 7 for a case which could be considered representative of random noise characteristics.

It became apparent that a truly useful solution in our initial application, the location of the internal standard, had to be capable of handling positively skewed distributions of different types with occasional non-unique modes. Complex EICP were investigated first.

Complex EICP can be viewed as consisting of a mixture of data and background noise intensities. The relative proportion of each indicates the degree of complexity for the m/z value in the data file in question. For the threshold to be established in complex data files, a mechanism is required to separate data from background noise. By examining lists of intensity distributions for background corrected EICP, it became apparent that discrete intensity values with a low frequency of occurrence were predominantly data. Even for EICP with regular spaced peaks, this trait was apparent as can be seen in Figure 8.

In Figure 8, a listing of discrete background intensity values and their frequency of occurrence for the entire EICP is represented. Two distinct groups of intensity value/frequency pairs can be observed in this list: high intensity/low frequency and low intensity/high frequency. Typically, any positive intensity value with a frequency of occurrence less than or equal to two was a value that we would have assigned to the data region in manual integration. This was noted with few exceptions for a variety of complex EICP, and subsequently the intensity frequency was utilized to separate background and data by eliminating intensity values (frequency filtering) from the background corrected intensity distribution with frequency $\leq 2$. It remained to determine the exact intensity level that would define the noise level in the frequency filtered distribution.

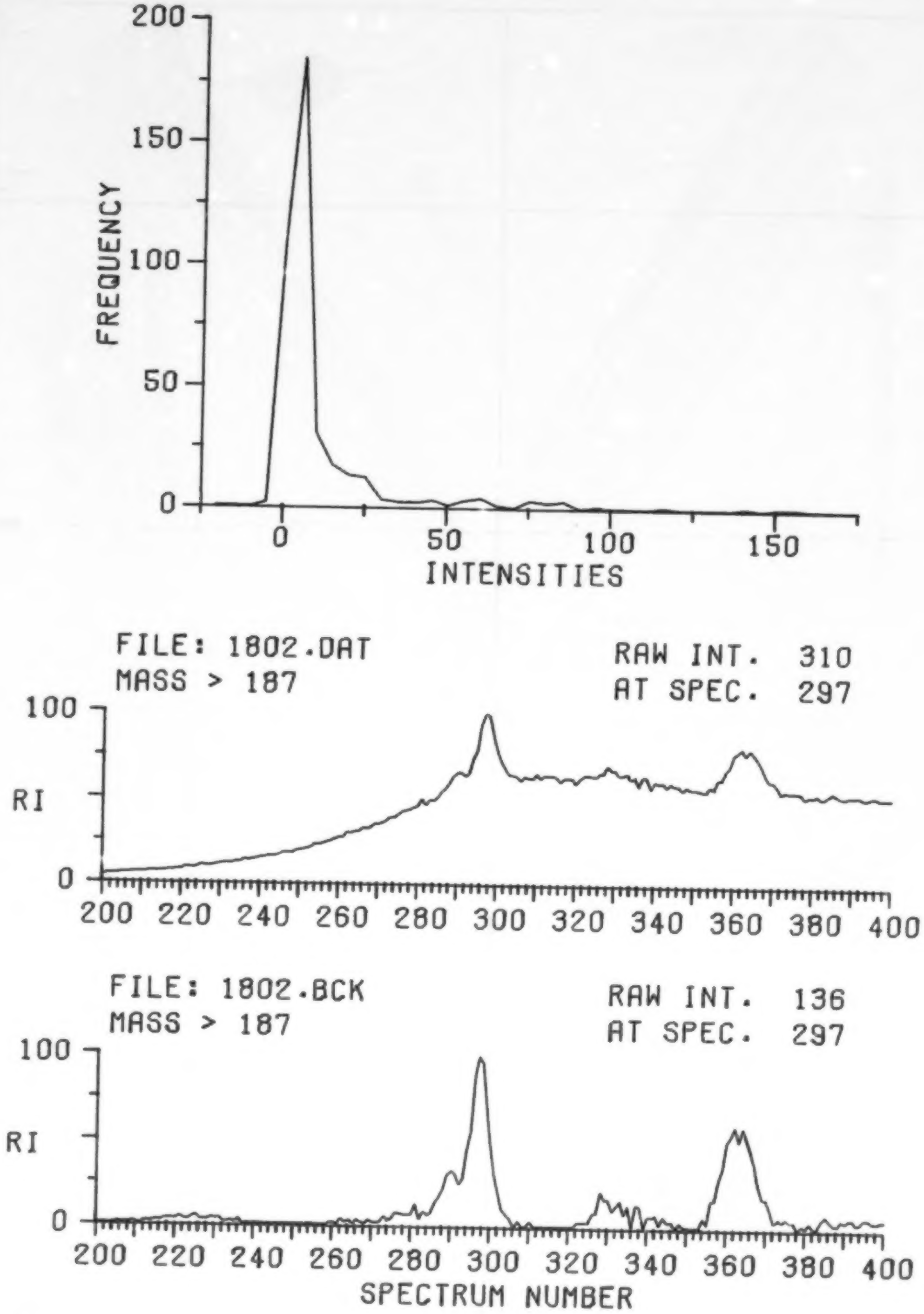


Figure 7. Plot of ranged intensities for normal distribution. Uncorrected and corrected data shown in lower plots.

```
                      INTENSITY / FREQUENCY
  -14.    1       21.    2       53.    2      148.    1
  -12.    1       22.    6       54.    1      149.    1
  -10.    1       23.    1       56.    1      153.    2
   -8.    1       24.    5       58.    1      154.    2
   -5.    1       25.    8       60.    1      162.    1
   -4.    1       26.    5       61.    2      164.    1
   -3.    1       27.    4       62.    2      172.    1
   -2.    3       28.    3       63.    2      176.    1
   -1.    6       29.    5       64.    1      177.    1
    0.  350       30.    9       68.    2      182.    2
    1.   28       31.    3       69.    1      192.    1
    2.    7       32.    7       70.    1      195.    1
    3.   17       34.    5       71.    2      198.    1
    4.   19       35.    9       79.    1      211.    1
    5.   12       36.    3       81.    1      215.    1
    6.   10       37.    2       83.    1      223.    1
    7.    7       38.    1       84.    2      248.    1
    8.    5       39.    4       88.    1      261.    1
    9.    3       40.    4       91.    1      262.    1
   10.    2       41.    2       94.    1      263.    1
   11.    1       42.    2      100.    1      292.    1
   12.    2       43.    5      101.    1      322.    1
   13.    7       44.    2      105.    1      335.    1
   14.    2       45.    1      107.    1      369.    1
   15.    6       46.    4      112.    1      372.    1
   16.   14       47.    2      134.    1      392.    1
   17.    1       48.    4      136.    1      482.    1
   18.    1       49.    1      137.    1
   19.    6       50.    2      139.    1
   20.    4       51.    2      147.    1
```

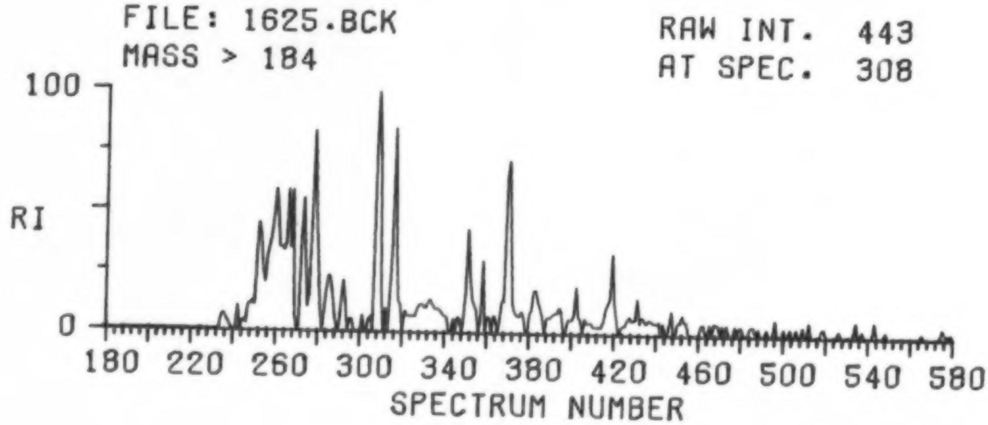


Figure 8. Background corrected EICP and frequency distribution for complex case.

Ideally this technique had to handle different distributions successfully because different distributions were observed in our data. For this reason, we decided to set the threshold independent of assumptions about the distribution type. By applying the EICP peak finding program to EICP with easily observed internal standards, minimum confusion was produced when the threshold was set at the 99th percentile of intensities in the frequency filtered background-corrected EICP. In the case of a Gaussian distribution, our procedure for threshold determination results in virtually identical results with that of equation (4); however, our approach is flexible because the distribution is not required to be of any one type.

Peak Detection. A rather simplistic peak detection algorithm is employed to point to the location of the sought components. Our experience has been that the conditions for internal standard location are rigorous enough, however, to preclude false positive identifications. The algorithm basically looks for a peak in the EICP that is above the threshold value. If such a feature is found, a second EICP characteristic of the component of interest is examined for a corresponding peak with maximum intensity ± 2 scans from the first peak maximum. The more EICP that are examined, the more rigorous is the identification procedure. If simultaneous peaking is not observed (± 2 scans), subsequent features in the EICP are examined until the end of the EICP is reached. Our procedure utilizes eight ions for identification of the internal standards. As few as two ions, provided they are judiciously chosen, have been shown to be effective for compound class recognition. The component location software is capable of examining the data for features in 30 EICP concurrently.

## RESULTS & DISCUSSION

Internal Standard Location. The automated GC/MS data survey system (CLEANUP/PBM)[9] in initial tests was unable to find the internal standard in about 50 percent of the data. The initial test of this approach, then, was to locate the internal standard in the mass of GC/MS data for each sample. The background correction and thresholding technique was applied to selected EICP belonging to the mass spectrum of the internal standard for the acid and base/neutral fractions, anthracene-$d_{10}$. Twelve data files from the same contract laboratory were chosen initially that used the same GC columns (1% SP-1240DA). Six of these files were examples of cases in which CLEANUP/PBM had failed to find the internal standard. Upon application of the preprocessor to all 12 data files, results were obtained that agreed with CLEANUP/PBM. For the data files in which no internal standard was located by either technique, manual examination of the files in question also found no internal standard.

It was hoped that this test would demonstrate that the preprocessor could locate the internal standard when CLEANUP/PBM had failed, but because these files did not have internal standards another test was required. At the same time, however, it is important to note that the preprocessor showed excellent resistance to false positive. As importantly, the ability to "flag" data files with no internal standard was demonstrated.

To demonstrate the ability of the approach to locate the internal standard where CLEANUP/PBM had failed, hundreds of such data files were run through the preprocessor. The preprocessor located anthracene-d10 in 90% of these data files in agreement with their manually assigned retention time, but 5% could not be found at all. The reason that the internal standard was not located in some cases by the preprocessor was apparently due to the low absolute intensity values of the low abundance, low mass fragments. As a result, peaks were not found for these masses and the preprocessor algorithms rejected these runs.

In Figure 9, possible pitfalls for automated identification of anthracene-d10 by specific mass search is shown. Masses 188 and 186 are both seen at scans 366 and 485. An examination of the relative intensities indicates that at both scans anthracene-d10 is indicated. Examination of the spectra shows that the peak at 485 is due to anthraquinone-d8 while anthracene-d10 is at 366. Thus, either retention data or additional masses are necessary to give a high probability identification. In

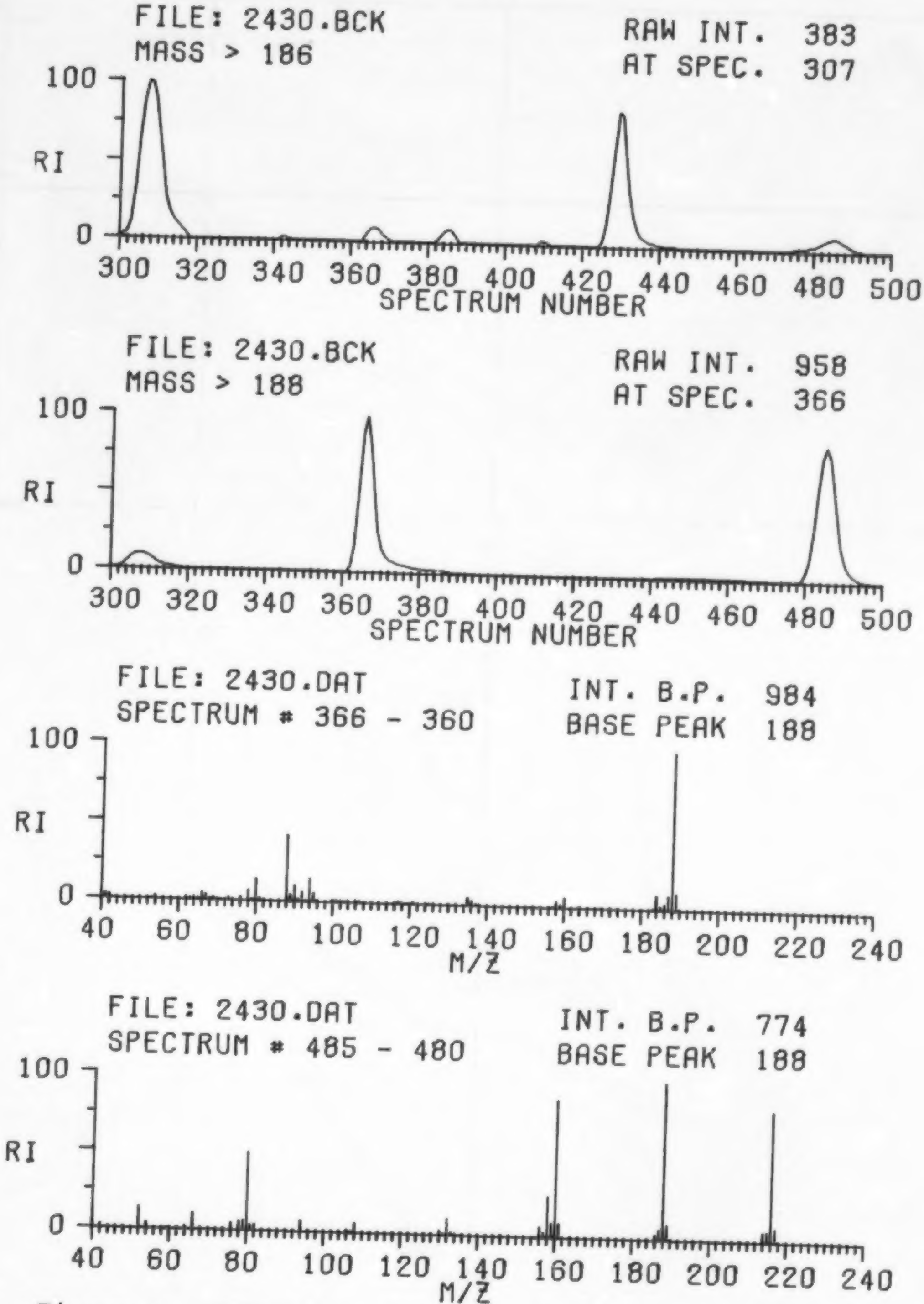


Figure 9. EICP plots for mass 188 and 186 with spectra for anthracene-$d_{10}$ and anthraquinone-$d_8$.

the case of finding an internal standard, only additional masses are appropriate when dealing with a wide variety of data sources.

Over the course of the project, the anthracene-d10 recognition rate increased approximately 90% to 95% overall by utilizing the preprocessor coupled with the CLEANUP/PBM modules in the system. In general, manual examination of the file in question is warranted only on about 5-10% of all data at this time.

In the purgeable fraction, both bromochloromethane and 1,4-dichlorobutane were used as internal standards. Both compounds are plagued by interferences in many sample matrices and present special problems that are not so readily noticeable when dealing with the highly recognizable spectrum of anthracene-d10. Problems arising in location of the internal standard bromochloromethane after background removal are shown in Figure 10. In the topmost frame m/z 49 is depicted. This ion is essential in the identification of bromochloromethane, but in this illustration its occurrence at spectrum 130 is dwarfed by the huge peak centered about spectrum 100. In fact use of the frequency derived threshold (mode + 2.5 $\sigma$ mode) eliminates the peak at 130. Use of the frequency filtering technique for determining threshold (99th percentile of the intensities occurring 3 or more times) preserves this peak in the presence of a much larger one. In the middle and lower frames of Figure 10, the importance of choosing internal standards that have little interference from sample components is shown. Mass 128 is the molecular ion for 1,4-dichlorobutane-$^{35}$Cl, $^{37}$Cl. The low abundance of m/z 128 requires that other ions much less singular be used for identification. Ions at m/z 90 and m/z 92 can be used in many cases because the isotopic abundance ratios (for $C_4H_7{}^{35}Cl^+$ and $C_4H_7{}^{37}Cl^+$) should give strong indications for detection. In Figure 10 it can be seen that m/z 128 is barely discernable at scan 326. However, m/z 90 and 92 are completely obscured by toluene and are of no use in this case. The most intense ion of the 1,4-dichlorobutane spectrum, 55, is of little use due to an extremely convoluted EICP as shown in Figure 3. Mass 55 may be seen in Figure 3. In this case the automated preprocessor gives negative results--as it should.

## Location of Chlorinated Biphenyls

Another application of a compound specific presearch was in detecting the presence of chlorinated biphenyls. A data file in which the CLEANUP/PBM programs had detected a tetrachlorobiphenyl species was selected for initial study. Utilizing the four most intense masses in a reference mass spectrum (m/z - 292, 220, 290, and 294), we applied background correction with intensity threshold determination and the EICP peak finding

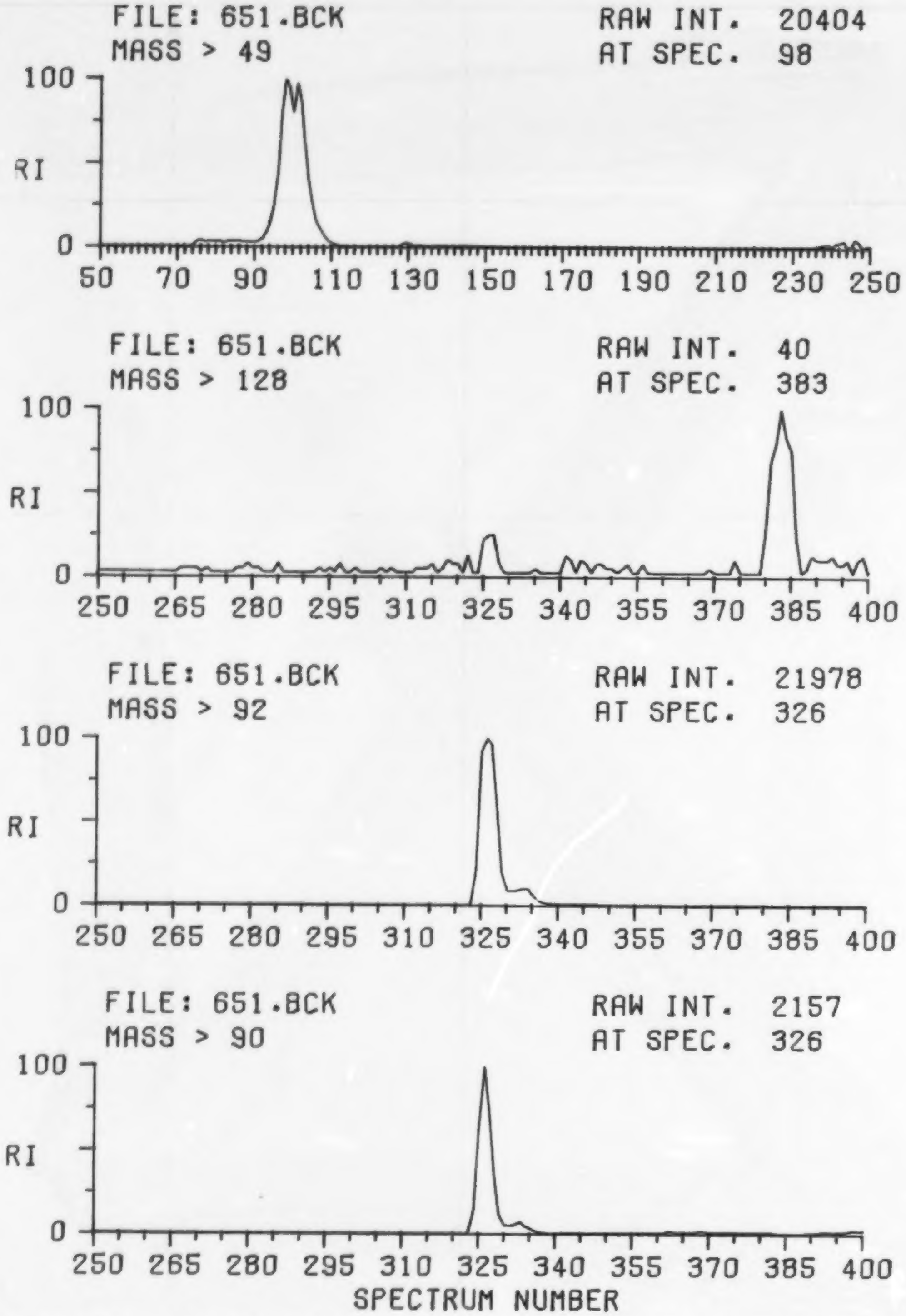


Figure 10. Illustration of problems in finding bromochloromethane (top) and 1,4-dichlorobutane (lower three plots).

routine to determine whether we could locate the same tetrachlorobiphenyl species that CLEANUP/PBM had found. Our approach located the compound at the spectrum number in agreement with CLEANUP/PBM. The background corrected EICP had relative intensities for the selected m/z values in good agreement with the standard reference spectrum matched by PBM as shown in Table 3.

TABLE 3. RELATIVE INTENSITY AGREEMENT BETWEEN BACKGROUND CORRECTED AND STANDARD REFERENCE SPECTRUM FOR SELECTED M/Z VALUES

| Tetrachlorobiphenyl Library Entry No. 13110 CAS No. 11097691 | | Sample Data Relative Intensities After Background Correction | |
|---|---|---|---|
| 292 | 100 | 292 | 100 |
| 220 | 83 | 220 | 86 |
| 290 | 79 | 290 | 77 |
| 294 | 48 | 294 | 53 |

Moreover, three additional eluants were detected in this data file with similar mass spectral characteristics at nearby retention times. Subsequent manual examination of these spectra verified the 4 and 2 chlorine cluster in the m/z 290 and 220 mass range. In summary, our approach agreed with CLEANUP/PBM for the location of the tetrachlorobiphenyl based on EICP of selected mass. As importantly, additional eluants with similar mass spectral properties to tetrachlorobiphenyl were indicated for further study. This application demonstrated the usefulness of the preprocessor in extracting selected scans in complex data files for manual study. A similar approach was employed to search 65 GC/MS data files for the presence of selected chlorinated biphenyls.

In this application, we sought to provide, where possible, GC/MS confirmation to GC (electron capture) results that had indicated the presence of chlorinated biphenyls in effluents of a given industrial category. To this end, 65 GC/MS data files (greater than 25,000 mass spectra) were preprocessed in the fashion outlined above for mono, di, tri, and tetrachlorobiphenyl. The peak finding algorithm required that three of four characteristic masses exhibit themselves in the thresholded, background corrected data within ±1 scan of each other. This presearch resulted in condensing the 25,000 mass spectra to 35 mass spectra. For all but two of these spectra, manual examination determined that chlorinated biphenyls were not present. In the two data files remaining, chlorinated aromatics were present but the eluants were not the chlorinated biphenyls. Manual study showed that these eluants were probable dichloro and trichloro nitrogen containing aromatic species.

All data files in question were also studied manually to confirm the preprocessor results. The results of manual examination (plotting EICP base peaks for selected m/z values for selected chlorinated biphenyls) agreed with the presearch in that the compounds searched for were not indicated in the data.

The usefulness of this technique in examining large volumes of GC/MS data is apparent. We do not mean to imply, however, that this approach should be employed to unequivocally determine presence/absence of organic compounds in any data file without prior knowledge of retention data. Nor do we intend to imply that this type of EICP presearch can replace more in depth manual or computer based data analysis for survey analysis. Rather, we have demonstrated that this simplistic approach can be used to condense large volumes of GC/MS data to a humanly manageable fraction of the total data analysis task. The risk involved in misassignment is dependent on the "uniqueness" of the m/z values selected, so that this technique should always be used with caution and with other confirmatory techniques. In practice, when "unique" ions are prevalent in the mass spectrum of interest, this approach can be employed in large volumes of GC/MS data to highlight spectra for further examination. The fact that this approach has been applied successfully to large volumes of GC/MS data generated by a variety of analytical laboratories utilizing different GC/MS instrumentation and chromatographic configurations is noteworthy.

# REFERENCES

1. Shackelford, W. M. and L. H. Keith. Evolution of the Priority Pollutant List from the Consent Decree. In: Proceedings of the Second Open Forum on Management of Petroleum Wastewater. EPA-600/2-78-058, U.S. Environmental Protection Agency, Ada, Oklahoma, 1978. pp. 103-109.

2. Keith, L. H. and W. A. Telliard. Priority Pollutants I - A Prospective View. Environ. Sci. Technol. 13(4):416-423, 1979.

3. Guidelines Establishing Test Procedures for the Analysis of Pollutants; Proposed Regulations, Federal Register, 44(233), Monday, December 3, 1979, 69464-69575.

4. Nau, H. and K. Biemann, Computer-Assisted Assignments of Retention Indices in Gas Chromatography-Mass Spectrometry and its Application to Mixtures of Biological Origin, Analytical Chemistry, 46(3):426-434, 1974.

5. Sweeley, C. C., N. D. Young, J. F. Holland, and S. C. Gates, Rapid Computerized Identification of Compounds in Complex Biological Mixtures by Gas Chromatography-Mass Spectrometry, Journal of Chromatography, 99:507-517, 1974.

6. Smith, Dennis H., Michael Achenbach, William J. Yaeger, Patricia J. Anderson, William L. Fitch, and Thomas C. Rindfleisch. Quantitative Comparison of Combined Gas Chromatography/Mass Spectrometric Profiles of Complex Mixtures. Anal. Chem., 49(11):1623-1632, 1977.

7. Tsuchiya, Yoshio, J. Boulanger, and K. Sumi. Identification of Chemical Compounds by Computer Matching of Low Resolution Mass Spectra and Retention Indices, Chromatographia, 10(3):154-156, 1977.

8. Blaisdell, B. Edwin. Automatic Computer Construction, Maintenance, and Use of Specialized Joint Libraries of Mass Spectra and Retention Indices from Gas Chromatography-Mass Spectrometry Systems, Analytical Chemistry, 49(1):180-186, 1977.

9. Shackelford, W. M., D. M. Cline, L. O. Burchfield, L. Faas, G. R. Kurth, and A. D. Sauter. A Computer Survey of GC/MS Data Acquired in EPA's Priority Pollutant Screening Analysis: System and Results, American Chemical Society National Meeting, San Francisco, California, August 27, 1980.

10. Summons, R. E., W. E. Periera, W. E. Reynolds, T. C. Rindfleisch, and A. M. Duffield. Analysis of Twelve Amino Acids in Biological Fluids by Mass Fragmentog- raphy. Anal. Chem., 46(4):582-587, 1974.

11. Eichelberger, J. W., Harris, L. E., and W. L. Budde. Reference Compound to Calibrate Ion Abundance Measurement in Gas Chromatography-Mass Spectrometry Systems, Analytical Chemistry, 47:995-1000, 1975.

NOTE: Mention of trade names or commercial products does not constitute endorsement or recommendation for use by the U.S. Environmental Protection Agency.